



Guide de EMR publication d'Amazon

Amazon EMR



Amazon EMR: Guide de EMR publication d'Amazon

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Les marques et la présentation commerciale d'Amazon ne peuvent être utilisées en relation avec un produit ou un service qui n'est pas d'Amazon, d'une manière susceptible de créer une confusion parmi les clients, ou d'une manière qui dénigre ou discrédite Amazon. Toutes les autres marques commerciales qui ne sont pas la propriété d'Amazon appartiennent à leurs propriétaires respectifs, qui peuvent ou non être affiliés ou connectés à Amazon, ou sponsorisés par Amazon.

Table of Contents

À propos d'Amazon EMR Releases	1
Support standard	2
Comprendre les EMR versions d'Amazon	2
Gestion des versions des versions	3
Politique de prise en charge	4
Considérations	7
Versions publiées par Amazon EMR 7.x	9
Versions des applications dans les versions EMR 7.x d'Amazon	10
emr-7.2.0	10
emr-7.1.0	66
emr-7.0.0	94
Versions publiées par Amazon EMR 6.x	132
Versions des applications dans les versions 6.x d'Amazon EMR	134
emr-6.15.0	134
emr-6.14.0	183
emr-6.13.0	232
emr-6.12.0	288
emr-6.11.1	349
emr-6.11.0	406
emr-6.10.1	444
emr-6.10.0	503
emr-6.9.1	545
emr-6.9.0	592
emr-6.8.1	641
emr-6.8.0	688
emr-6.7.0	736
emr-6.6.0	796
emr-6.5.0	859
emr-6.4.0	888
emr-6.3.1	923
emr-6.3.0	951
emr-6.2.1	985
emr-6.2.0	1013
emr-6.1.1	1047

emr-6.1.0	1068
emr-6.0.1	1095
emr-6.0.0	1113
Versions publiées par Amazon EMR 5.x	1135
Versions des applications dans les versions 5.x d'Amazon EMR	1138
emr-5.36.2	1139
emr-5.36.1	1166
emr-5.36.0	1213
emr-5.35.0	1251
emr-5.34.0	1279
emr-5.33.1	1305
emr-5.33.0	1335
emr-5.32.1	1358
emr-5.32.0	1383
emr-5.31.1	1412
emr-5.31.0	1431
emr-5.30.2	1455
emr-5.30.1	1475
emr-5.30.0	1497
emr-5.29.0	1521
emr-5.28.1	1540
emr-5.28.0	1559
emr-5.27.1	1579
emr-5.27.0	1596
emr-5.26.0	1616
emr-5.25.0	1636
emr-5.24.1	1656
emr-5.24.0	1673
emr-5.23.1	1692
emr-5.23.0	1709
emr-5.22.0	1728
emr-5.21.2	1748
emr-5.21.1	1764
emr-5.21.0	1782
emr-5.20.1	1801
emr-5.20.0	1818

emr-5.19.1	1838
emr-5.19.0	1855
emr-5.18.1	1873
emr-5.18.0	1890
emr-5.17.2	1907
emr-5.17.1	1924
emr-5.17.0	1940
emr-5.16.1	1958
emr-5.16.0	1974
emr-5.15.1	1991
emr-5.15.0	2007
emr-5.14.2	2024
emr-5.14.1	2040
emr-5.14.0	2056
emr-5.13.1	2074
emr-5.13.0	2090
emr-5.12.3	2106
emr-5.12.2	2122
emr-5.12.1	2137
emr-5.12.0	2153
emr-5.11.4	2170
emr-5.11.3	2186
emr-5.11.2	2201
emr-5.11.1	2217
emr-5.11.0	2233
emr-5.10.1	2249
emr-5.10.0	2264
emr-5.9.1	2281
emr-5.9.0	2296
emr-5.8.3	2313
emr-5.8.2	2328
emr-5.8.1	2343
emr-5.8.0	2358
emr-5.7.1	2375
emr-5.7.0	2390
emr-5.6.1	2405

emr-5.6.0	2420
emr-5.5.4	2436
emr-5.5.3	2451
emr-5.5.2	2466
emr-5.5.1	2481
emr-5.5.0	2496
emr-5.4.1	2512
emr-5.4.0	2527
emr-5.3.2	2542
emr-5.3.1	2557
emr-5.3.0	2572
emr-5.2.3	2587
emr-5.2.2	2602
emr-5.2.1	2617
emr-5.2.0	2632
emr-5.1.1	2647
emr-5.1.0	2662
emr-5.0.3	2677
emr-5.0.2	2691
emr-5.0.1	2705
emr-5.0.0	2719
Versions publiées par Amazon EMR 4.x	2734
Versions des applications dans les versions 4.x d'Amazon EMR	2736
Différences de version	2736
emr-4.9.6	2784
emr-4.9.5	2799
emr-4.9.4	2813
emr-4.9.3	2827
emr-4.9.2	2842
emr-4.9.1	2856
emr-4.8.5	2871
emr-4.8.4	2885
emr-4.8.3	2899
emr-4.8.2	2914
emr-4.8.1	2928
emr-4.8.0	2942

emr-4.7.4	2957
emr-4.7.3	2971
emr-4.7.2	2984
emr-4.7.1	2998
emr-4.7.0	3011
emr-4.6.1	3026
emr-4.6.0	3038
emr-4.5.0	3052
emr-4.4.0	3063
emr-4.3.0	3076
emr-4.2.0	3087
emr-4.1.0	3097
emr-4.0.0	3106
Versions 2.x et AMI 3.x	3113
Création d'un cluster	3115
Installation des applications	3117
Personnalisation des configurations	3117
Hive	3125
HBase	3135
Pig	3149
Spark	3155
S3 DistCp	3158
Nouveautés	3162
Amazon EMR 7.2.0	3162
Amazon EMR 6.15.0	3165
Amazon EMR 5.36.2	3184
Compatibilité SigV4	3187
Approche d'atténuation CVE -2021-44228	3188
Solution d'action Amazon EMR Bootstrap pour Log4j CVE -2021-44228 et -2021-45046	
CVE	3189
Questions fréquentes (FAQ)	3196
Archivage	3199
Version 6.14.0	3199
Version 6.13.0	3219
Version 6.12.0	3244
Version 6.11.1	3274

Version 6.11.0	3301
Version 6.10.0	3312
Version 6.9.0	3325
Version 6.8.0	3344
Version 6.7.0	3365
Version 6.6.0	3398
Version 5.35.0	3434
Version 5.34.0	3438
Version 6.5.0	3441
Version 6.4.0	3443
Version 5.32.0	3452
Version 6.2.0	3457
Version 5.31.0	3466
Version 6.1.0	3471
Version 6.0.0	3478
Version 5.30.1	3483
Version 5.30.0	3488
Version 5.29.0	3494
Version 5.28.1	3495
Version 5.28.0	3497
Version 5.27.0	3499
Version 5.26.0	3501
Version 5.25.0	3504
Version 5.24.1	3507
Version 5.24.0	3509
Version	3511
Version 5.22.0	3513
Version 5.21.1	3516
Version 5.21.0	3518
Version 5.20.0	3520
Version 5.19.0	3524
Version 5.18.0	3526
Version 5.17.1	3527
Version 5.17.0	3527
Version 5.16.0	3528
Version 5.15.0	3530

Version 5.14.1	3531
Version 5.14.0	3531
Version 5.13.0	3534
Version 5.12.2	3534
Version 5.12.1	3534
Version 5.12.0	3535
Version 5.11.3	3537
Version 5.11.2	3537
Version 5.11.1	3537
Version 5.11.0	3538
Version 5.10.0	3539
Version 5.9.0	3541
Version 5.8.2	3543
Version 5.8.1	3543
Version 5.8.0	3544
Version 5.7.0	3546
Version 5.6.0	3547
Version 5.5.3	3547
Version 5.5.2	3548
Version 5.5.1	3548
Version 5.5.0	3548
Version 5.4.0	3550
Version 5.3.1	3551
Version 5.3.0	3551
Version 5.2.2	3552
Version 5.2.1	3552
Version 5.2.0	3553
Version 5.1.0	3554
Version 5.0.3	3555
Version 5.0.0	3555
Version 4.9.5	3557
Version 4.9.4	3557
Version 4.9.3	3557
Version 4.9.2	3558
Version 4.9.1	3558
Version 4.8.4	3559

Version 4.8.3	3559
Version 4.8.2	3560
Version 4.8.0	3560
Version 4.7.2	3561
Version 4.7.1	3562
Version 4.7.0	3562
Version 4.6.0	3564
Version 4.5.0	3566
Version 4.4.0	3566
Version 4.3.0	3568
Version 4.2.0	3570
Configuration des applications	3572
Configuration des applications lorsque vous créez un cluster	3575
Fournir une configuration dans la console lorsque vous créez un cluster	3575
Fournissez une configuration à l'aide du AWS CLI lorsque vous créez un cluster	3575
Fournissez une configuration à l'aide de Java SDK lorsque vous créez un cluster	3576
Reconfigurer un groupe d'instances dans un cluster en cours d'exécution	3577
Considérations à prendre en compte lors de la reconfiguration d'un groupe d'instances	3578
Reconfigurer un groupe d'instances dans la console	3581
Reconfigurez un groupe d'instances à l'aide du CLI	3582
Reconfigurer un groupe d'instances à l'aide de Java SDK	3587
Dépannage	3588
Stocker les données de configuration sensibles dans AWS Secrets Manager	3591
Créer un secret	3592
Accordez EMR à Amazon l'accès pour récupérer le secret	3592
Utiliser le secret dans une classification de configuration	3592
Mettre à jour la valeur d'un secret	3594
Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique	3594
Considérations	3594
Remplacez le JVM	3597
Ports de service	3599
Utilisateurs de l'application	3601
Vérification des dépendances à l'aide du référentiel d'artefact	3602
EMR Système de fichiers (EMRFS)	3605
Vue cohérente	3607
Activation de la vue cohérente	3611

Comprendre comment EMRFS une vue cohérente permet de suivre les objets dans Amazon S3	3613
Logique des nouvelles tentatives	3614
EMRFSmétadonnées de vue cohérentes	3616
Configurer les notifications de cohérence pour CloudWatch et Amazon SQS	3619
Configuration de la vue cohérente	3621
EMRFSCLIRéférence de commande	3626
Autoriser l'accès aux EMRFS données dans Amazon S3	3637
Création d'un fournisseur d'informations d'identification personnalisé pour EMRFS les données dans Amazon S3	3638
Gestion du point de AWS Security Token Service terminaison par défaut	3640
Spécification du chiffrement Amazon S3 à l'aide EMRFS des propriétés	3641
Utilisation AWS KMS keys pour le EMRFS chiffrement	3642
Chiffrement côté serveur sur Amazon S3	3643
Chiffrement côté client sur Amazon S3	3646
CloudWatch Agent Amazon	3654
Créer un cluster	3656
Autorisations nécessaires	3656
Point de terminaison requis	3657
Créer le cluster	3657
Métriques par défaut	3658
Configuration	3661
Amazon EMR 7.1.0	3661
Amazon EMR 7.0.0	3696
Considérations	3702
Historique	3702
Delta Lake	3704
Introduction	3704
Utilisation des clusters de Delta Lake	3705
Delta Lake avec Flink	3705
Delta Lake avec Trino	3710
Delta Lake avec Spark	3711
Delta Lake avec Spark et Glue	3718
Considérations	3718
Historique	3719
Flink	3721

Création d'un cluster avec Flink	3724
Configuration de Flink	3725
Hive et Glue	3725
Fichier de configuration	3727
Plusieurs nœuds primaires	3729
Taille du processus de mémoire	3730
Taille du fichier de sortie du journal	3730
Java 11	3732
Tâches Flink	3736
Démarez une application Flink YARN en tant qu'étape sur un cluster de longue durée	3737
Soumettre du travail à une application Flink existante sur un cluster de longue durée	3739
Soumission d'une tâche Flink transitoire	3740
Shell Flink Scala	3742
Interface utilisateur Flink	3743
Outil de mise à l'échelle automatique Flink	3744
Présentation	3744
Considérations	3745
Activation de l'outil de mise à l'échelle automatique	3745
Configurations	3747
Optimisation des temps de redémarrage	3753
Récupération locale des tâches	3753
Points de contrôle incrémentiels	3754
Récupération précise	3755
Redémarrage combiné	3756
Flink avec Zeppelin	3757
Introduction	3757
Prérequis	3757
Configurer Zeppelin-Flink sur un cluster EMR	3758
Exécuter des tâches Flink avec Zeppelin-Flink sur un cluster EMR	3759
Historique des publications de Flink	3764
Notes de mise à jour de Flink par version	3802
Ganglia	3803
Création d'un cluster à l'aide de Ganglia	3805
Affichage des métriques Ganglia	3806
Métriques Hadoop et Spark dans Ganglia	3807
Historique des versions de Ganglia	3808

Hadoop	3854
Configuration de Hadoop	3856
Configuration de la tâche	3856
Paramètres de configuration de démon Hadoop	4217
HDFSconfiguration	4487
Chiffrement transparent HDFS sur Amazon EMR	4488
Configuration du chiffrement HDFS transparent	4489
Considérations relatives au chiffrement HDFS transparent	4492
Serveur de gestion des clés Hadoop	4492
HDFSchiffrement transparent sur les EMR clusters comportant plusieurs nœuds principaux	4496
Création ou exécution d'une application Hadoop	4498
Créer des fichiers binaires à l'aide d'Amazon EMR	4498
Traitement des données avec Streaming	4501
Traitez les données avec un outil personnalisé JAR	4507
Lire les objets restaurés	4510
Exemples	4511
Considérations	4512
Activer la détection non uniforme de l'accès à la mémoire pour les conteneurs YARN	4512
Historique des versions de Hadoop	4514
Notes de mise à jour de Hadoop par version	4576
HBase	4582
Création d'un cluster avec HBase	4586
Création d'un cluster à HBase l'aide de la console	4586
Création d'un cluster à HBase l'aide du AWS CLI	4586
HBasesur Amazon S3 (mode de stockage Amazon S3)	4587
Activation HBase sur Amazon S3	4589
Utilisation d'un cluster réplica en lecture	4590
HFileSuivi permanent	4592
Considérations opérationnelles	4594
Journaux d'écriture anticipée (WAL) pour Amazon EMR	4598
WALespaces de travail	4600
Autorisations nécessaires	4601
Activant WAL	4601
Restauration à partir de WAL	4604
Configurations de la sécurité	4605

En utilisant AWS PrivateLink	4606
WALprix et indicateurs	4607
Marquage WAL des espaces de travail	4609
Considérations et disponibilité	4610
EMRWALCLIréférence	4612
Utilisation de la HBase coque	4615
Créer une table	4616
Placement d'une valeur	4616
Obtention d'une valeur	4616
Supprimer une table	4616
Accédez aux HBase tables avec Hive	4616
Utilisation de HBase snapshots	4618
Création d'un instantané à l'aide d'une table	4618
Suppression d'un instantané	4619
Affichage des informations d'instantané	4619
Exportation d'un instantané vers Amazon S3	4619
Importer un instantané depuis Amazon S3	4620
Restaurer une table à partir d'instantanés dans le shell HBase	4621
Configurez HBase	4622
Modifications apportées à l'allocation de mémoire dans YARN	4623
HBasenuméros de port	4624
HBaseparamètres du site pour optimiser	4624
Afficher l'interface HBase utilisateur	4626
Afficher les fichiers journaux HBase	4628
Moniteur HBase avec Ganglia	4629
Migration depuis les versions précédentes HBase	4631
HBasehistorique des versions	4631
HCatalog	4703
Création d'un cluster avec HCatalog	4704
En utilisant HCatalog	4705
Désactiver l'écriture directe lors de l'utilisation HCatalog HStorer	4706
Créez une table en utilisant HCat CLI et utilisez ces données dans Pig	4706
Accès à la table à l'aide de Spark SQL	4707
Exemple : créer une HCatalog table et y écrire en utilisant Pig	4709
HCataloghistorique des versions	4710
Hive	4779

Différences et considérations relatives à Hive sur Amazon EMR	4782
Différences entre Apache Hive sur Amazon EMR et Apache Hive	4782
Différences dans Hive entre les EMR versions 4.x et 5.x d'Amazon	4783
Fonctionnalités supplémentaires de Hive sur Amazon EMR	4784
Configuration d'un métastore externe pour Hive	4789
Utiliser le catalogue de données AWS Glue comme métastore pour Hive	4790
Utilisation d'une base de SQL données My Database externe ou d'Amazon Aurora	4797
Utilisation du pilote Hive JDBC	4800
Amélioration des performances de Hive	4802
Activation du validateur optimisé Hive EMRFS S3	4803
Utilisation de S3 Select	4804
MSCKOptimisation	4806
Utiliser Hive LLAP	4807
Activer LLAP	4808
Commencez par LLAP votre cluster	4809
Vérifier le LLAP statut	4810
Démarrer ou arrêter LLAP	4810
Redimensionner le nombre de LLAP démons	4810
Chiffrement dans Hive	4810
Chiffrement modulaire du parquet dans Hive	4811
Chiffrement en transit dans HS2	4815
Historique des versions de Hive	4816
Notes de mise à jour de Hive par version	4886
Hudi	4972
Comment fonctionne Hudi	4974
Comprendre les types de stockage des jeux de données : copie en écriture ou fusion en lecture	4975
Enregistrement d'un jeu de données Hudi auprès de votre métastore	4976
Considérations et restrictions	4976
Créer un cluster avec Hudi installé	4978
Travailler avec un jeu de données Hudi	4979
Initialisation d'une session Spark pour Hudi	4983
Écrire dans un jeu de données Hudi	4983
Données d'Upsert	4988
Supprimer un enregistrement	4990
Lire à partir d'un jeu de données Hudi	4991

Utilisez le Hudi CLI	4993
Historique des sorties de Hudi	4994
Hue	4997
Versions de Hue	4997
Fonctionnalités prises en charge et non prises en charge de Hue sur Amazon EMR	4999
Considérations	5000
Performances avec de grandes tables de métadonnées Hue	5001
Incompatibilité entre les versions de Hue	5002
Connexion à l'interface utilisateur Web de Hue	5002
Utilisation de Hue avec une base de données distante sur Amazon RDS	5003
Résolution des problèmes	5005
Configurations avancées pour Hue	5006
Configurer Hue pour les LDAP utilisateurs	5006
Historique des versions de Hue	5009
Iceberg	5072
Comment fonctionne Iceberg	5073
Utiliser un cluster avec Iceberg	5075
Utiliser un cluster Iceberg avec Spark	5075
Utiliser un cluster Iceberg avec Trino	5080
Utiliser un cluster Iceberg avec Flink	5082
Utiliser un cluster Iceberg avec Hive	5087
Considérations et restrictions	5090
Considérations relatives à l'utilisation d'Iceberg avec Spark	5090
Considérations relatives à l'utilisation d'Iceberg avec Trino	5091
Considérations relatives à l'utilisation d'Iceberg avec Flink	5091
Considérations relatives à l'utilisation d'Iceberg avec Hive	5091
Historique des versions d'Iceberg	5092
Notes de mise à jour d'Iceberg par version	5093
Bloc-notes Jupyter	5096
EMR Studio	5096
Bloc-notes EMR	5096
JupyterHub	5097
Créez un cluster avec JupyterHub	5101
Considérations à prendre en compte lors de l'utilisation JupyterHub sur Amazon EMR	5103
Configuration JupyterHub	5104
Configuration de la persistance pour les blocs-notes dans Amazon S3	5105

Connexion au nœud principal et aux serveurs de bloc-notes	5106
JupyterHub configuration et administration	5107
Ajout d'utilisateurs et d'administrateurs de bloc-notes Jupyter	5109
Installation de noyaux et de bibliothèques supplémentaires	5121
JupyterHub historique des versions	5125
Livy	5162
Activant HTTPS	5164
Historique des versions de Livy	5165
MXNet	5209
MXNethistorique des versions	5211
Oozie	5240
Utiliser Oozie avec une base de données distante sur Amazon RDS	5242
Configurer la version Java pour Oozie	5245
Historique des versions d'Oozie	5246
Notes de mise à jour d'Oozie par version	5303
Phoenix	5304
Création d'un cluster avec Phoenix	5307
Personnalisation des configurations Phoenix	5308
Clients Phoenix	5309
Historique des sorties de Phoenix	5312
Pig	5387
Soumission de travail Pig	5389
Soumettez le travail de Pig à l'aide de la EMR console Amazon	5390
Soumettez le travail de Pig à l'aide du AWS CLI	5391
Appel de fonctions définies par l'utilisateur à partir de Pig	5392
JARFichiers d'appels depuis Pig	5393
Appel de scripts Python/Jython à partir de Pig	5393
Historique de versions Pig	5394
Presto et Trino	5460
Utilisation de Presto avec le catalogue de données AWS Glue	5463
Spécifier le catalogue de données AWS Glue comme métastore	5464
IAMautorisations	4794
Considérations relatives à l'utilisation du catalogue de données AWS Glue	5469
Utilisation de S3 Select Pushdown	5469
S3 Select Pushdown est-il adapté à mon application ?	5470
Considérations et restrictions	5470

Activation de S3 Select Pushdown avec PrestoDB ou Trino	5471
Ajout de connecteurs de base de données	5472
En utilisantSSL/TLS et LDAPS	5473
Utilisation de l'LDAP authentication	5474
Activation du mode strict Presto	5482
Considérations	5484
Gestion de la perte d'instances Spot dans Presto	5484
Exécution tolérante aux pannes	5487
Configuration	5487
Responsable des échanges	5487
Considérations et restrictions	5489
Utilisation du dimensionnement automatique de Presto avec désaffectation gracieuse	5489
Considérations relatives à Presto sur Amazon EMR	5491
Exécutable en ligne de commande Presto	5491
Propriétés de déploiement Presto non configurables	5491
Installation de PrestoDB et Trino	5492
EMRFS configuration de Presto3 FileSystem	5493
Paramètre par défaut de l'emprunt d'identité de l'utilisateur final	5494
Port par défaut de l'interface Web Presto	5494
Problème d'exécution du compartiment Hive dans certaines versions	5494
Historique des versions de Presto	5495
Notes de mise à jour de Trino (PrestoSQL) par version	5560
Spark	5563
Création d'un cluster Spark	5567
Exécutez des applications Spark avec Docker sur Amazon 6.x EMR	5569
Considérations	5570
Création d'une image Docker	5571
Utilisation d'images Docker provenant d'Amazon ECR	5572
Utiliser le catalogue de données AWS Glue comme métastore pour Spark SQL	5577
Spécifier le catalogue de données AWS Glue comme métastore	5578
IAM autorisations	4794
Considérations	4795
Configuration de Spark	5583
Paramètres par défaut de Spark	5584
Configuration du récupérateur de mémoire	5586
maximizeResourceAllocation	5587

Comportement de mise hors service	5589
Variable d' ThriftServer environnement Spark	5592
Modification des paramètres Spark par défaut	5593
Migration d'Apache Log4j 1.x vers Log4j 2.x	5595
Optimisation des performances de Spark	5596
Exécution de requêtes adaptative	5596
Nettoyage dynamique de partition	5598
Aplatissement des sous-requêtes scalaires	5601
DISTINCTavant INTERSECT	5602
Jonction de filtre Bloom	5603
Réorganisation optimisée des jonctions	5603
Mise en cache des fragments de résultats	5604
Activation de la mise en cache des fragments de résultats Spark	5605
Considérations	5606
Utiliser l'RAPIDSaccélérateur	5608
Choix des types d'instance	5608
Configuration des applications	5608
Ajout d'une action d'amorçage	5616
Lancement de votre cluster	5617
Accès au shell de Spark	5617
Utiliser Amazon SageMaker Spark pour le machine learning	5619
Ecriture d'une application Spark	5619
Scala	5619
Java	5620
Python	5621
Amélioration des performances de Spark avec S3	5623
Utiliser S3 Select	5623
EMRFSÉmetteur optimisé pour S3	5627
Utilisez le protocole de EMRFS validation optimisé pour S3	5635
Nouvelle tentative pour les demandes S3	5642
Ajout d'une étape Spark	5645
Remplacement des paramètres de configuration par défaut de Spark	5648
Afficher l'historique de l'application Spark	5649
Accédez au site Web Spark UIs	5649
Utilisation de Spark avec Amazon Kinesis Data Streams	5650
Utilisation de Spark sur Amazon Redshift	5650

Lancement d'une application Spark	5651
Authentification dans Amazon Redshift	5652
Lecture et écriture vers Amazon Redshift	5654
Considérations	5656
Historique des versions de Spark	5657
Sqoop	5721
Versions de Sqoop	5721
Considérations relatives à Sqoop sur Amazon EMR	5723
Utilisation de Sqoop avec intégration HCatalog	5723
Support de Sqoop JDBC et de base de données	5723
Sécurisation de votre mot de passe	5725
Historique des versions de Sqoop	5726
TensorFlow	5779
TensorFlow compilations par type d'EC2instance Amazon	5781
Sécurité	5781
En utilisant TensorBoard	5781
TensorFlow historique des versions	5782
Tez	5805
Création d'un cluster avec Tez	5807
Configuration de Tez	5808
Exemple de configuration	5808
Ouverture fractionnée asynchrone Tez	5809
Interface utilisateur web Tez	5811
Serveur de chronologie	5812
Historique des versions de Tez	5812
Notes de publication de Tez par version	5856
Zeppelin	5864
Considérations relatives à l'utilisation de Zeppelin sur Amazon EMR	5866
Historique des versions de Zeppelin	5867
ZooKeeper	5925
ZooKeeper historique des versions	5927
Connecteurs et utilitaires	5966
Exportation, interrogation et jointure de tables dans DynamoDB	5966
Configuration d'une table Hive pour exécuter des commandes Hive	5968
Exemples de commandes Hive pour l'exportation, l'importation et l'interrogation de données	5977

Optimisation des performances	5987
Kinesis	5991
Que puis-je faire avec l'intégration entre Amazon EMR et Amazon Kinesis ?	5991
Analyse de points de contrôle des flux Amazon Kinesis	5991
Considérations sur les performances	5993
Planifiez une analyse Amazon Kinesis avec Amazon EMR	5993
Migration du connecteur Spark Kinesis vers la version 2.x SDK pour Amazon 7.0 EMR	5993
S3 DistCp (s3-dist-cp)	6003
DistCp Options S3	6004
Ajouter S3 DistCp en tant qu'étape dans un cluster	6012
Nettoyage après l'échec des DistCp tâches S3	6014
Exécuter des commandes et des scripts sur un cluster	6016
Soumettre une étape JAR personnalisée pour exécuter un script ou une commande	6017
Autres moyens d'utiliser <code>command-runner.jar</code>	6018
Glossaire AWS	6020
.....	6021

À propos d'Amazon EMR Releases

Une EMR version d'Amazon est un ensemble d'applications open source issues de l'écosystème des mégadonnées. Chaque version comprend différentes applications, composants et fonctionnalités de Big Data que vous sélectionnez pour qu'Amazon EMR installe et configure lorsque vous créez un cluster. Les applications sont empaquetées à l'aide d'un système basé sur [Apache BigTop](#), un projet open source associé à l'écosystème Hadoop. Ce guide fournit des informations sur les applications incluses dans les EMR versions d'Amazon.

Pour plus d'informations sur la prise en main et l'utilisation d'AmazonEMR, consultez le [guide EMR de gestion Amazon](#).

Lorsque vous lancez un cluster, vous pouvez choisir parmi plusieurs versions d'AmazonEMR. Vous pouvez ainsi tester et utiliser des versions d'applications qui répondent à vos exigences de compatibilité. Vous spécifiez le numéro de version avec l'étiquette de version. Les étiquettes de version sont sous la forme `emr-x.x.x`. Par exemple, `emr-7.2.0`.

À partir d'Amazon EMR 5.18.0, vous pouvez utiliser le référentiel d'EMRartefacts Amazon pour créer votre code de travail en fonction des versions exactes des bibliothèques et des dépendances disponibles avec des versions spécifiques d'Amazon. EMR Pour plus d'informations, consultez [Vérification des dépendances à l'aide du référentiel d'artefacts d'Amazon EMR](#).

Pour recevoir des mises à jour lorsqu'une nouvelle EMR version d'Amazon est disponible, abonnez-vous au [RSSflux des notes de EMR publication d'Amazon](#).

Informations sur les dernières versions, y compris les versions des applications, les notes de version, les composants et les classifications de configuration des EMR séries Amazon 7.x, 6.x et 5.x :

- [Amazon EMR version 7.2.0](#)
- [Amazon EMR version 6.15.0](#)
- [Amazon EMR version 5.36.2](#)

Note

Les nouvelles EMR versions d'Amazon sont mises à disposition dans différentes régions sur une période de plusieurs jours, en commençant par la première région à la date de

sortie initiale. Il est possible que la dernière version ne soit pas disponible dans votre région pendant cette période.

Notes de mise à jour concernant les dernières EMR versions d'Amazon et historique de toutes les versions :

- [Nouveautés](#)
- [EMR Archive des notes de publication sur Amazon](#)

Un historique complet des versions des applications dans chaque EMR version d'Amazon :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Détails de chaque EMR version d'Amazon et différences entre les séries de versions, le cas échéant :

- [Versions publiées par Amazon EMR 7.x](#)
- [Versions publiées par Amazon EMR 6.x](#)
- [Versions publiées par Amazon EMR 5.x](#)
- [Versions publiées par Amazon EMR 4.x](#)
- [Versions Amazon EMR 2.x et AMI 3.x](#)

Support EMR standard d'Amazon

Comprendre les EMR versions d'Amazon

La EMR version Amazon est le véhicule qui fournit les logiciels nécessaires pour exécuter vos applications open source sur des EC2 plateformes EKS ou sans serveur. Une EMR version d'Amazon est composée de trois éléments : l'environnement d'exécution, les moteurs principaux et les extras.

- Environnement d'exécution : l'environnement d'exécution inclut le système d'exploitation inclus dans Amazon Machine Image (AMI) ou dans une autre image de conteneur lorsque vous lancez votre cluster ou conteneur (par exemple, Amazon Linux 2023). Il inclut également les environnements d'exécution de langage tels que Amazon JDK Corretto, ainsi que d'autres outils connexes installés par l'image.
- Moteurs principaux : il s'agit notamment des principaux logiciels open source tels qu'Apache Hive et des formats de table ouverts tels qu'Apache Iceberg.
- Extras — Il s'agit notamment de bibliothèques pratiques et de packages Python tels que `mariaadb-connector-java` des logiciels open source tels qu'Apache Pig.

La EMR version de lancement d'Amazon que vous sélectionnez regroupe tous ces composants dans une nouvelle version d'Amazon Machine Images () ou d'images de conteneur basées sur Amazon Linux (ALAMI). Amazon EMR gère la même version AL pour tous les nœuds d'un cluster, d'un EKS conteneur ou d'une application sans serveur. Amazon EMR vise à obtenir le dernier environnement d'exécution dans les 90 jours suivant leur publication par Amazon Linux, et les nouvelles versions open source de Core Engines dans les 90 jours suivant leur sortie en amont. Les extras sont publiés selon les besoins. Vous trouverez la liste des logiciels dans les notes de publication de chaque EMR version d'Amazon.

Gestion des versions des versions

Pour vous aider à comprendre l'ampleur des modifications apportées à chaque version, Amazon EMR utilise le versionnement sémantique. Le versionnement sémantique suit un format de `<major>.<minor>.<patch>` Prenons l'exemple de la version 7.0.0.

Le premier chiffre indique la version majeure, ou version majeure. Les versions majeures apportent généralement des modifications, des améliorations et de nouvelles fonctionnalités substantielles à l'environnement d'exécution ou au moteur principal qui ne sont pas rétrocompatibles. Par exemple, la version majeure 7.x utilise Amazon Linux 2023 avec Amazon JDK Corretto 17 par défaut. Par conséquent, cette version contient plusieurs modifications majeures par rapport à la version majeure 6.x, qui est livrée sur Amazon Linux 2 et avec Amazon JDK Corretto 8 par défaut.

Le deuxième chiffre indique la version mineure, ou version mineure. Les versions mineures sont des versions non rétrocompatibles qui contiennent des modifications, des améliorations et des fonctionnalités incrémentielles apportées aux moteurs principaux et aux extras, ainsi que de nouvelles fonctionnalités. Par exemple, la version 6.15 utilise Apache Spark 3.4.1, alors que la version 6.11.1 est fournie avec Apache Spark 3.3.2.

Le troisième chiffre indique une version de correctif, ou une version de correctif. Les versions de correctif sont destinées à être des versions rétrocompatibles contenant des correctifs et des mises à jour pour Core Engine et les options supplémentaires, mais aucune nouvelle fonctionnalité ou mise à niveau de OSS version. Par exemple, les versions 6.11.1 et 6.11.0 contiennent toutes deux la même version d'Apache Spark 3.3.2. OSS Depuis la version 6.6.x, la dernière version de correctif propose le dernier correctif d'environnement d'exécution disponible. L'environnement d'exécution n'affecte pas la publication du correctif d'AmazonEMR. Par exemple, 6.11.1 a été lancé avec la version 2.0.20240109.0 d'AL du 19 décembre 2023 au 10 janvier 2024, mais avec la version AL 2.0.20240124.0 du 11 janvier 2024. Certaines EKS versions d'Amazon EMR on vous permettent d'ajouter des balises ou des suffixes aux versions sémantiques afin de vous proposer des fonctionnalités alternatives. Par exemple, la `emr-6.15.0-latest` est sortie d'Amazon EMR le EKS lance une version compatible avec Amazon Corretto JDK 8, tandis que la `emr-6.15.0-java17-latest` est sortie d'EMR Amazon EKS vous permet de lancer une version compatible avec Amazon Corretto 17. JDK

Politique de prise en charge

Introduction

Amazon EMR a pour objectif de publier de nouvelles versions mineures au moins une fois tous les 90 jours et de fournir une assistance pour les versions mineures pendant une période de 24 mois à compter de la date de sortie. Ce support couvre l'environnement d'exécution et les moteurs principaux avec leurs dépendances associées, mais ne couvre pas les extras ni leurs dépendances associées. L'environnement d'exécution inclut également des composants de APIs plate-forme propriétaires nécessaires au démarrage, à l'arrêt et au fonctionnement des clusters et des applications. Ce calendrier de publication cohérent garantit un cycle prévisible, ce qui vous permet de planifier, de tester et de passer plus facilement à une version offrant support et sécurité.

À quoi s'attendre avec le Support standard

Le support standard fournit des correctifs sur les tickets de support technique pour les problèmes que vous rencontrez avec les composants Runtime Environment et Core Engines dans le cadre des configurations recommandées. Tous les correctifs sont soumis à disponibilité. La configuration recommandée est l'utilisation d'Amazon EMR sans aucune modification, ajout ou modification des fichiers binaires et des configurations présents dans une EMR version d'Amazon, à l'exception de ceux figurant dans la [EMRdocumentation Amazon](#). Amazon EMR déploie des correctifs sur le dernier correctif, la version mineure ou majeure de la version d'Amazon EMR dans les 90 jours suivant la vérification du correctif par nos soins. Amazon applique EMR automatiquement des correctifs lorsque

vous lancez un nouveau conteneur Amazon EMR onEC2, lorsque vous lancez un nouveau EKS conteneur Amazon EMR on ou que vous déclenchez une nouvelle tâche EMR sans serveur. Les composants supplémentaires sont fournis pour des raisons de commodité et Amazon EMR ne fournit pas de correctifs liés aux suppléments.

Composants

Le support standard couvre les composants de l'environnement d'exécution et des moteurs principaux, par exemple le système d'exploitation, les environnements d'exécution des langages et les principaux logiciels open source tels qu'Apache Hive et Apache Iceberg. Vous trouverez la liste complète des composants pris en charge pour chaque version dans [À propos d'Amazon EMR Releases](#) [Amazon Linux FAQs](#).

La liste suivante décrit le support que nous fournissons pour les différents types de composants dans le cadre du support standard :

- Composants de l'environnement d'exécution : les composants de l'environnement d'exécution recevront des correctifs sur les tickets de support technique. Les correctifs sont classés comme suit : (a) bogues critiques, (b) problèmes critiques de corruption de données et (c) problèmes de sécurité critiques. Si vous êtes éligible, Amazon EMR rétroportera les correctifs pour Runtime Environment vers les anciennes versions. Pour des raisons de compatibilité opérationnelle avec les composants open source du moteur principal, certains composants de l'environnement d'exécution doivent conserver des versions spécifiques afin de ne pas endommager les applications des clients. Pour ces composants, Amazon dépend EMR de l'open source en amont pour la disponibilité des correctifs. Lorsque des correctifs seront disponibles en open source, nous fournirons la dernière version stable dans les 90 jours suivant leur vérification par AmazonEMR.
- Composants de Core Engines : Core Engines fournit les dernières versions pour de nombreux projets open source, chacun contenant des centaines de bibliothèques de dépendances transitives. Alors que les communautés open source qui gèrent ces projets tentent fréquemment de résoudre les problèmes et de résoudre les vulnérabilités et expositions courantes connues (CVE), les dernières versions peuvent toujours contenir des bogues connus et. CVEs Amazon EMR dépend de l'open source en amont pour la disponibilité des correctifs et fournira la dernière version stable dans le cadre des composants du moteur principal dans les 90 jours suivant leur vérification par AmazonEMR. Dans certains cas, Amazon EMR peut fournir un correctif pour un CVE composant du moteur principal qui doit être traité avant l'open source en amont. Amazon vous fournit EMR également un support technique et des correctifs pour les fonctionnalités ajoutées en plus de l'open source dans les configurations recommandées. Nous ne rétroportons pas les correctifs pour les composants du moteur principal vers d'anciennes versions de correctifs ou de versions mineures.

- Composants supplémentaires : Amazon EMR ne prend pas en charge les composants supplémentaires. Les composants supplémentaires sont des projets open source fournis pour des raisons de commodité et Amazon EMR ne fournit pas de solutions aux problèmes rencontrés avec eux. Toute demande d'assistance ou correction peut être traitée par le biais de la communauté open source supportant ces composants.

Support standard : cycle de vie

Ce qui suit décrit les étapes du cycle de vie du Support standard :

- Support standard : les EMR versions Amazon sont éligibles au support standard 24 mois à compter de la date de sortie. Vous pouvez créer des tickets de support technique et vous attendre à des mises à jour pour les problèmes que vous rencontrez avec ces versions.
- Fin du support : une fois le support standard terminé, les EMR versions Amazon entrent dans la phase de fin de support (eOS) pendant 12 mois. Les versions d'eOS ne sont pas éligibles au support technique et vous ne pourrez pas créer de tickets pour les clusters, les conteneurs ou les jobs exécutés sur ces versions. Les versions d'EoS ne recevront aucun correctif, correctif ou mise à jour. Les versions d'eOS seront supprimées de la console, mais continueront d'être disponibles via API et AWS CLI. Vous pouvez toujours continuer à exécuter des charges de travail sur les versions d'eOS. Nous vous recommandons vivement de migrer vers la dernière EMR version d'Amazon afin de continuer à recevoir les correctifs de sécurité, de rester éligible au support technique et de créer des tickets d'assistance en cas de besoin.
- Fin de vie : après la fin de la période de Support (EoS), les versions sont considérées comme étant en fin de vie (EoL). Bien que vous puissiez continuer à utiliser des clusters EoL, Amazon EMR se réserve le droit de supprimer les versions EoL du API et SDK sur une case-by-case base déterminée pour des raisons de sécurité et d'exploitation. Nous vous recommandons vivement de migrer vers la dernière version d'AmazonEMR, car les versions d'EoL peuvent être supprimées du API et SDK dans des cas exceptionnels.

Support de pont

Amazon EMR a annoncé cette nouvelle politique de support le 25 juillet 2024. En vertu de cette politique, les versions d'Amazon EMR publiées le 24 juillet 2022 ou avant cette date sont désormais désignées comme étant en fin de support. Toutefois, afin de vous donner plus de temps pour planifier et migrer vers des versions plus récentes, Amazon EMR proposera un Bridge Support, équivalent au support standard, pour ces anciennes versions publiées dans les deux ans précédant cette annonce.

Après le 25 juillet 2024, vous pouvez consulter le statut actuel et les délais de support dans les notes de publication.

Le tableau suivant indique l'état de support de toutes les EMR versions Amazon existantes au moment de l'annonce de la politique, le 25 juillet 2024 :

Versions et périodes prises en charge

EMR Version publiée par Amazon	Date de sortie initiale	Date de fin du support standard	Date de début de fin du support	Date de début de fin de vie
7.2.0	25 juillet 2024	24 juillet 2026	25 juillet 2026	25 juillet 2027
7.1.0	23 avril 2024	22 avril 2026	23 avril 2026	23 avril 2027
7.0.0	19 décembre 2023	18 décembre 2025	19 décembre 2025	19 décembre 2026
5.36.x et 6.6.x — 6.15.x	9 mai 2022 au 13 novembre 2023	Support du pont jusqu'au 24 janvier 2026	25 janvier 2026	25 janvier 2026
<ul style="list-style-type: none"> Série 6.x : 6.5.0 et versions antérieures Série 5.x : 5.35.0 et versions inférieures Séries 4.x, 3.x et 2.x 	1er janvier 2013 au 30 mars 2022	Support du pont jusqu'au 24 juillet 2025	25 juillet 2025	25 juillet 2025

Considérations

Le support standard est disponible pour tous les modèles de EMR déploiement Amazon (EMREMRsur EC2EKS, Amazon sur et EMR sans serveur), dans toutes les régions où Amazon

EMR est disponible, sans frais supplémentaires. Les clusters exécutés avec des configurations recommandées sont automatiquement éligibles au support tel que décrit dans la politique. Vous n'avez donc pas besoin de prendre d'autres mesures pour activer le support.

- Le support standard prend uniquement en charge les composants requis pour les EMR clusters Amazon. Amazon ne peut pas garantir les correctifs de sécurité et la disponibilité des correctifs dans le cas où les composants open source de Core Engine atteignent la fin de vie en amont, ou lorsque les mises à jour de sécurité ne sont plus disponibles pour les dépendances. Bien que vous puissiez choisir d'installer des Extras, Amazon ne les prendra pas en charge ni EMR ne prendra en charge leurs dépendances. Par exemple, vous pouvez installer des applications tierces dans votre environnement personnalisé AMI pour renforcer la sécurité de votre cluster, installer des composants supplémentaires ou copier des objets à l'aide de scripts d'action bootstrap, ou SSH dans votre cluster et mettre à niveau les versions de package par défaut. Amazon EMR ne prend pas en charge ces composants. Le support standard ne couvre pas les actions de démarrage, les packages, les bibliothèques fournis par le client, votre code bring-your-own personnalisé et les applications personnalisées que vous pouvez configurer Amazon pour installer EMR à votre convenance.
- Vos clusters existants ne seront pas affectés, quelle que soit la EMR version d'Amazon qu'ils utilisent. Vous pouvez continuer à exécuter les clusters existants sans interruption. Vous pouvez également continuer à lancer de nouveaux clusters et à exécuter des tâches sur toutes les versions existantes ou nouvelles. Toutes les versions existantes et les nouvelles versions au moment de l'entrée en vigueur de la politique sont couvertes par le Support Standard pendant 24 mois à compter de la date de sortie initiale de la version de EMR lancement d'Amazon. Amazon vous fournira une assistance temporaire lors de l'annonce initiale de la politique. Pour bénéficier d'une assistance ininterrompue, nous vous conseillons de tester vos applications et de les mettre à niveau rapidement vers la EMR version la plus récente d'Amazon.
- Amazon EMR ne modifiera pas les composants du Support Standard sur les versions ou clusters existants. Amazon EMR se réserve toutefois le droit de respecter la fin de vie en amont sur une case-by-case base déterminée et de supprimer ces composants dans les nouvelles versions, ou de passer la version existante au statut de fin de support (eOS) ou de fin de vie (eOL) dans des cas exceptionnels. Nous vous informerons de toute suppression par le biais des canaux disponibles.
- Lorsqu'une nouvelle région est lancée, Amazon EMR ne prend en charge que les EMR versions publiées par Amazon dans le cadre du support standard, publiées six mois avant la date de mise à disposition générale de la nouvelle région.

- Amazon EMR ne mettra pas automatiquement à jour vos clusters existants vers les dernières versions. Toutefois, vous pouvez choisir de mettre à jour les nouveaux clusters avec les dernières versions des correctifs si vous le souhaitez.
- Le coût de votre EMR utilisation non prise en charge d'Amazon sera pris en compte dans votre AWS facture. Même si vous utilisez Amazon de manière non prise EMR en charge, les coûts associés à cette utilisation font toujours partie de votre AWS consommation globale et seront inclus dans le calcul de vos frais de support.

Pour plus d'informations, contactez le [Support aux AWS développeurs](#).

Versions publiées par Amazon EMR 7.x

Cette section contient les versions des applications, les notes de mise à jour, les versions des composants et les classifications de configuration disponibles dans chaque version d'Amazon EMR 7.x.

Lorsque vous lancez un cluster, vous pouvez choisir parmi plusieurs versions d'AmazonEMR. Vous pouvez ainsi tester et utiliser des versions d'applications qui répondent à vos exigences de compatibilité. Vous spécifiez le numéro de version avec l'étiquette de version. Les étiquettes de version sont sous la forme `emr-x.x.x`. Par exemple, `emr-7.2.0`.

Les nouvelles EMR versions d'Amazon sont mises à disposition dans différentes régions sur une période de plusieurs jours, en commençant par la première région à la date de sortie initiale. Il est possible que la dernière version ne soit pas disponible dans votre région pendant cette période.

Pour un tableau complet des versions des applications dans chaque version EMR 7.x d'Amazon, consultez [Versions des applications dans les versions EMR 7.x d'Amazon](#).

Rubriques

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Amazon EMR version 7.2.0](#)
- [Amazon EMR version 7.1.0](#)
- [Amazon EMR version 7.0.0](#)

Versions des applications dans les versions EMR 7.x d'Amazon

Pour obtenir un tableau complet répertoriant les versions d'applications disponibles dans chaque version d'Amazon EMR 7.x, ouvrez les versions des [applications dans les versions d'Amazon EMR 7.x](#) dans votre navigateur.

Amazon EMR version 7.2.0

Cycle de vie pris en charge en 7.2.0

Le tableau suivant décrit les dates de cycle de vie prises en charge pour la EMR version 7.2.0 d'Amazon.

Phase de support	Date	
Date de sortie initiale	25 juillet 2024	
Support standard jusqu'à	24 juillet 2026	
Fin de l'assistance	25 juillet 2026	
Fin de vie	25 juillet 2027	

Pour obtenir une liste des vulnérabilités et des risques courants critiques et élevés (CVEs) qui n'affectent pas les EMR clusters dans le cadre des logiciels et des configurations recommandés, voir [7.2.0 Vulnérabilités et expositions courantes connues pour les moteurs principaux](#). Pour une liste des problèmes CVEs corrigés dans la version 7.2.0, voir [7.2.0 a corrigé les vulnérabilités et expositions courantes](#).

7.2.0 vulnérabilités et risques courants

Cette page répertorie toutes les vulnérabilités et expositions courantes (CVEs) liées à Amazon EMR 7.2.0.

7.2.0 vulnérabilités et risques courants connus pour les moteurs principaux

Le tableau suivant répertorie tout CVEs ce qui n'a aucun impact sur les EMR clusters exécutés sur Amazon EMR 7.2.0. Amazon EMR dépend de l'open source en amont pour la disponibilité des correctifs et fournira la dernière version stable dans le cadre des composants du moteur principal dans les 90 jours suivant la EMR vérification des correctifs par Amazon.

CVEID	Sévérité	CVE détails URL
CVE-2024-1597	Critique	https://nvd.nist.gov/vuln/detail/CVE-2024-1597
CVE-2023-44981	Critique	https://nvd.nist.gov/vuln/detail/CVE-2023-44981
CVE-2023-31047	Critique	https://nvd.nist.gov/vuln/detail/CVE-2023-31047
CVE-2023-25668	Critique	https://nvd.nist.gov/vuln/detail/CVE-2023-25668
CVE-2023-25664	Critique	https://nvd.nist.gov/vuln/detail/CVE-2023-25664
CVE-2_46337	Critique	https://nvd.nist.gov/vuln/detail/CVE-2_46337
CVE-2_42889	Critique	https://nvd.nist.gov/vuln/detail/CVE-42889
CVE-2_39135	Critique	https://nvd.nist.gov/vuln/detail/CVE-2_39135
CVE-2_37865	Critique	https://nvd.nist.gov/vuln/detail/CVE-2_37865
CVE-2_26612	Critique	https://nvd.nist.gov/vuln/detail/CVE-2_26612
CVE-2_25168	Critique	https://nvd.nist.gov/vuln/detail/CVE-2_25168
CVE-2_1471	Critique	https://nvd.nist.gov/vuln/detail/CVE-2_1471

CVEID	Sévérité	CVEdétails URL
CVE-2021-37404	Critique	https://nvd.nist.gov/vuln/detail/CVE-2021-37404
CVE-2021-35958	Critique	https://nvd.nist.gov/vuln/detail/CVE-2021-35958
CVE-2021-32798	Critique	https://nvd.nist.gov/vuln/detail/CVE-2021-32798
CVE-2021-32797	Critique	https://nvd.nist.gov/vuln/detail/CVE-2021-32797
CVE-2020-9548	Critique	https://nvd.nist.gov/vuln/detail/CVE-2020-9548
CVE-2020-9547	Critique	https://nvd.nist.gov/vuln/detail/CVE-2020-9547
CVE-2020-9546	Critique	https://nvd.nist.gov/vuln/detail/CVE-2020-9546
CVE-2020-8840	Critique	https://nvd.nist.gov/vuln/detail/CVE-2020-8840
CVE-2019-20445	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-20445
CVE-2019-20444	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-20444
CVE-2019-20330	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-20330
CVE-2019-17531	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-17531
CVE-2019-17267	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-17267

CVEID	Sévérité	CVEdétails URL
CVE-2019-17195	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-17195
CVE-2019-16943	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-16943
CVE-2019-16942	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-16942
CVE-2019-16335	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-16335
CVE-2019-14893	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-14893
CVE-2019-14892	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-14892
CVE-2019-14540	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-14540
CVE-2019-14379	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-14379
CVE-2019-10202	Critique	https://nvd.nist.gov/vuln/detail/CVE-2019-10202
CVE-2018-7489	Critique	https://nvd.nist.gov/vuln/detail/CVE-2018-7489
CVE-2018-19362	Critique	https://nvd.nist.gov/vuln/detail/CVE-2018-19362
CVE-2018-19361	Critique	https://nvd.nist.gov/vuln/detail/CVE-2018-19361
CVE-2018-19360	Critique	https://nvd.nist.gov/vuln/detail/CVE-2018-19360

CVEID	Sévérité	CVEdétails URL
CVE-2018-14721	Critique	https://nvd.nist.gov/vuln/detail/CVE-2018-14721
CVE-2018-14720	Critique	https://nvd.nist.gov/vuln/detail/CVE-2018-14720
CVE-2018-14719	Critique	https://nvd.nist.gov/vuln/detail/CVE-2018-14719
CVE-2018-14718	Critique	https://nvd.nist.gov/vuln/detail/CVE-2018-14718
CVE-2018-11307	Critique	https://nvd.nist.gov/vuln/detail/CVE-2018-11307
CVE-2017-7658	Critique	https://nvd.nist.gov/vuln/detail/CVE-2017-7658
CVE-2017-7657	Critique	https://nvd.nist.gov/vuln/detail/CVE-2017-7657
CVE-2017-7525	Critique	https://nvd.nist.gov/vuln/detail/CVE-2017-7525
CVE-2017-17485	Critique	https://nvd.nist.gov/vuln/detail/CVE-2017-17485
CVE-2017-15095	Critique	https://nvd.nist.gov/vuln/detail/CVE-2017-15095
CVE-2017-12629	Critique	https://nvd.nist.gov/vuln/detail/CVE-2017-12629
CVE-2016-5018	Critique	https://nvd.nist.gov/vuln/detail/CVE-2016-5018
CVE-2015-1832	Critique	https://nvd.nist.gov/vuln/detail/CVE-2015-1832

CVEID	Sévérité	CVEdétails URL
CVE-2024-29133	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2024-29133
CVE-2024-29131	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2024-29131
CVE-2024-24680	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2024-24680
CVE-2024-21634	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2024-21634
CVE-2023-6481	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6481
CVE-2023-6378	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-6378
CVE-2023-52428	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-52428
CVE-2023-4759	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-4759
CVE-2023-46695	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-46695
CVE-2023-46120	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-46120
CVE-2023-43665	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43665
CVE-2023-43642	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-43642
CVE-2023-41164	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-41164

CVEID	Sévérité	CVEdétails URL
CVE-2023-36478	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-36478
CVE-2023-36053	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-36053
CVE-2023-34610	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34610
CVE-2023-34478	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34478
CVE-2023-34455	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34455
CVE-2023-34454	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34454
CVE-2023-34453	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-34453
CVE-2023-30608	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-30608
CVE-2023-2976	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-2976
CVE-2023-27579	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-27579
CVE-2023-25801	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25801
CVE-2023-25676	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25676
CVE-2023-25675	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25675

CVEID	Sévérité	CVEdétails URL
CVE-2023-25674	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25674
CVE-2023-25673	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25673
CVE-2023-25672	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25672
CVE-2023-25671	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25671
CVE-2023-25670	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25670
CVE-2023-25669	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25669
CVE-2023-25667	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25667
CVE-2023-25666	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25666
CVE-2023-25665	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25665
CVE-2023-25663	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25663
CVE-2023-25662	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25662
CVE-2023-25660	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25660
CVE-2023-25659	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25659

CVEID	Sévérité	CVEdétails URL
CVE-2023-25658	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25658
CVE-2023-25194	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-25194
CVE-2023-24816	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-24816
CVE-2023-24580	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-24580
CVE-2023-23969	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-23969
CVE-2023-20883	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-20883
CVE-2023-1436	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1436
CVE-2023-1370	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-1370
CVE-2_46751	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_46751
CVE-2_45693	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_45693
CVE-2_45685	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_45685
CVE-2_44729	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_44729
CVE-2_42969	HIGH	https://nvd.nist.gov/vuln/detail/CVE-42969

CVEID	Sévérité	CVEdétails URL
CVE-2_42890	HIGH	https://nvd.nist.gov/vuln/detail/CVE -42890
CVE-avril 2004	HIGH	https://nvd.nist.gov/vuln/detail/CVE -04/2004
CVE-4 avril 2003	HIGH	https://nvd.nist.gov/vuln/detail/CVE -04.2003
CVE-2_41704	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_41704
CVE-2_40899	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_40899
CVE-2_40664	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_40664
CVE-2_40152	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_40152
CVE-2_40150	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_40150
CVE-2_40149	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_40149
CVE-2_40146	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_40146
CVE-2_37866	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_37866
CVE-2_36364	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_36364
CVE-2_3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2_3510

CVEID	Sévérité	CVEdétails URL
CVE-2_3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_3509
CVE-2_34169	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_34169
CVE-2_32532	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_32532
CVE-2_3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_3171
CVE-2_25647	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_25647
CVE-2_24758	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2_24758
CVE-2021-41303	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-41303
CVE-2021-37137	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37137
CVE-2021-37136	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-37136
CVE-2021-34538	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-34538
CVE-2021-33813	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-33813
CVE-2021-33036	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-33036
CVE-2021-31684	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-31684

CVEID	Sévérité	CVEdétails URL
CVE-2021-28165	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-28165
CVE-2021-26919	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-26919
CVE-2021-25646	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-25646
CVE-2020-9492	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-9492
CVE-2020-9480	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-9480
CVE-2020-7692	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-7692
CVE-2020-36518	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36518
CVE-2020-36189	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36189
CVE-2020-36188	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36188
CVE-2020-36187	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36187
CVE-2020-36186	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36186
CVE-2020-36185	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36185
CVE-2020-36184	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36184

CVEID	Sévérité	CVEdétails URL
CVE-2020-36183	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36183
CVE-2020-36182	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36182
CVE-2020-36181	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36181
CVE-2020-36180	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36180
CVE-2020-36179	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-36179
CVE-2020-35728	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35728
CVE-2020-35491	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35491
CVE-2020-35490	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35490
CVE-2020-35214	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35214
CVE-2020-35213	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-35213
CVE-2020-25649	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-25649
CVE-2020-24750	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24750
CVE-2020-24616	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-24616

CVEID	Sévérité	CVEdétails URL
CVE-2020-17533	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-17533
CVE-2020-17523	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-17523
CVE-2020-14195	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14195
CVE-2020-14062	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14062
CVE-2020-14061	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14061
CVE-2020-14060	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-14060
CVE-2020-13949	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-13949
CVE-2020-13936	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-13936
CVE-2020-11988	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11988
CVE-2020-11620	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11620
CVE-2020-11619	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11619
CVE-2020-11113	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11113
CVE-2020-11112	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11112

CVEID	Sévérité	CVEdétails URL
CVE-2020-11111	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-11111
CVE-2020-10969	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10969
CVE-2020-10968	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10968
CVE-2020-10673	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10673
CVE-2020-10672	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10672
CVE-2020-10650	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2020-10650
CVE-2019-17566	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-17566
CVE-2019-16869	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-16869
CVE-2019-14439	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-14439
CVE-2019-12086	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-12086
CVE-2019-10172	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-10172
CVE-2019-10099	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-10099
CVE-2019-02-05	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-0205

CVEID	Sévérité	CVEdétails URL
CVE-2019-02-04	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2019-0204
CVE-2018-8768	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-8768
CVE-2018-8012	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-8012
CVE-2018-5968	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-5968
CVE-2018-3827	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-3827
CVE-2018-3258	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-3258
CVE-2018-17190	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-17190
CVE-2018-12023	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-12023
CVE-2018-12022	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-12022
CVE-2018-11804	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-11804
CVE-2018-11793	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-11793
CVE-2018-10936	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-10936
CVE-2018-10054	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-10054

CVEID	Sévérité	CVEdétails URL
CVE-2017-9790	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-9790
CVE-2017-9735	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-9735
CVE-2017-7687	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-7687
CVE-2017-7656	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-7656
CVE-2017-5637	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-5637
CVE-2017-18640	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-18640
CVE-2017-15288	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-15288
CVE-2017-12612	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-12612
CVE-2017-1000034	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2017-1000034
CVE-2016-6796	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-6796
CVE-2016-1000352	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-1000352
CVE-2016-1000344	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-1000344
CVE-2016-1000343	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2016-1000343

CVEID	Sévérité	CVEdétails URL
CVE-2016-1000342	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2016-1000342
CVE-2016-1000340	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2016-1000340
CVE-2016-1000338	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2016-1000338
CVE-2015-7501	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2015-7501
CVE-2015-6420	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2015-6420
CVE-2015-5237	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2015-5237
CVE-2015-4852	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2015-4852
CVE-2015-2156	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2015-2156
CVE-2015-2080	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2015-2080
CVE-2014-01-14	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2014-0114
CVE-2013-4002	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2013-4002
CVE-2012-0881	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2012-0881

7.2.0 a corrigé des vulnérabilités et des risques courants pour les moteurs principaux

Le tableau suivant répertorie toutes les corrections apportées CVEs dans Amazon EMR 7.2.0.

CVEID	Sévérité	CVE détails URL
CVE-2019-12402	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-12402
CVE-2021-41561	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-41561

composants de la version 7.2.0

Amazon EMR 7.2.0 prend en charge les applications suivantes :

Moteurs principaux :

AmazonCloudWatchAgent, Delta, Flink, Hadoop, HiveHBase, HudiHCatalog, Livy, Iceberg, Phoenix, Spark, Tez, Trino JupyterEnterpriseGateway, gardien de zoo

Suppléments :

Hue JupyterHub, OozieMXNet, Pig, Presto, Sqoop, Zeppelin TensorFlow

Pour plus de détails sur les packages pris en charge, consultez les [détails des composants de la version 7.2.0](#). Pour plus d'informations sur les moteurs principaux et les options supplémentaires, consultez [Amazon EMR Standard Support](#).

Détails des composants de la version 7.2.0

Consultez le tableau suivant pour plus d'informations sur les packages Extras d'Amazon EMR 7.2.0.

Application	Type de demande	Package	Type de package	Version du package
AmazonCloudWatchAgent	Principal	hadoop-hdfs-journalnode	Suppléments	3.3.6. amzn.4-1. amzn2023

Application	Type de demande	Package	Type de package	Version du package
AmazonCloudWatchAgent	Principal	ranger-km	Suppléments	2.0.0-1 mai 2023
AmazonCloudWatchAgent	Principal	s3-dist-cp	Suppléments	2.32.0-1 amzn2023
AmazonCloudWatchAgent	Principal	gardien de zoo	Suppléments	3.9.1. amzn.1-1. amzn2023
AmazonCloudWatchAgent	Principal	zookeeper-server	Suppléments	3.9.1. amzn.1-1. amzn2023
Flink	Principal	hadoop-client	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-hdfs	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-hdfs-datanode	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-hdfs-journalnode	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-hdfs-namenode	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-httfs	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	Hadoop KMS	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-mapreduce	Suppléments	3.3.6. amzn.4-1. amzn2023

Application	Type de demande	Package	Type de package	Version du package
Flink	Principal	hadoop-mapreduce-historyserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	fil Hadoop	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-yarn-nodemanager	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-yarn-proxyserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-yarn-resourcemanager	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hadoop-yarn-timelineserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Flink	Principal	hudi	Suppléments	0.14.1.amzn.1-1.amzn2023
Flink	Principal	ranger-km	Suppléments	2.0.0-1 mai 2023
Flink	Principal	hadoop	Suppléments	3.3.6. amzn.4-1. amzn2023
HBase	Principal	s3-dist-cp	Suppléments	2.32.0-1 amzn2023
HCatalog	Principal	hadoop-yarn-proxyserver	Suppléments	3.3.6. amzn.4-1. amzn2023

Application	Type de demande	Package	Type de package	Version du package
HCatalog	Principal	mariadb-connector-java	Suppléments	2.7.2-1
Hive	Principal	hadoop-ftpfs	Suppléments	3.3.6. amzn.4-1. amzn2023
Hive	Principal	hadoop-yarn-proxyserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Hive	Principal	hive-hbase	Suppléments	3.1.3. amzn.11-1. amzn2023
Hive	Principal	mariadb-connector-java	Suppléments	2.7.2-1
Hive	Principal	s3-dist-cp	Suppléments	2.32.0-1 amzn2023
Phoenix	Principal	s3-dist-cp	Suppléments	2.32.0-1 amzn2023
Spark	Principal	noyau de données Spark	Suppléments	3.5.1. amzn.0-1. amzn2023
Tez	Principal	hadoop-mapreduce	Suppléments	3.3.6. amzn.4-1. amzn2023
Tez	Principal	hadoop-mapreduce-historyserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Tez	Principal	hadoop-yarn-proxyserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	bigtop-utils	Suppléments	1.2.0-1. amzn2023

Application	Type de demande	Package	Type de package	Version du package
Trino	Principal	hadoop-client	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-hdfs	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-hdfs-datanode	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-hdfs-journalnode	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-hdfs-namenode	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	Hadoop KMS	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-mapreduce	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-mapreduce-historyserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	fil Hadoop	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-yarn-nodemanager	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-yarn-proxyserver	Suppléments	3.3.6. amzn.4-1. amzn2023

Application	Type de demande	Package	Type de package	Version du package
Trino	Principal	hadoop-yarn-resourcemanager	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-yarn-timelineserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hive	Suppléments	3.1.3. amzn.11-1 . amzn2023
Trino	Principal	hive-hcatalog-server	Suppléments	3.1.3. amzn.11-1 . amzn2023
Trino	Principal	mariadb-connector-java	Suppléments	2.7.2-1
Trino	Principal	ranger-km	Suppléments	2.0.0-1 mai 2023
Trino	Principal	gardien de zoo	Suppléments	3.9.1. amzn.1-1. amzn2023
Trino	Principal	zookeeper-server	Suppléments	3.9.1. amzn.1-1. amzn2023
Trino	Principal	bigtop-groovy	Suppléments	2.5.4-1. amzn2023
Trino	Principal	bigtop-jsvc	Suppléments	1.2.4-1. amzn2023
Trino	Principal	hadoop	Suppléments	3.3.6. amzn.4-1. amzn2023
Trino	Principal	hadoop-lzo	Suppléments	0.4.19-1 amzn2023

Application	Type de demande	Package	Type de package	Version du package
Trino	Principal	catalogue hive-h	Suppléments	3.1.3. amzn.11-1 . amzn2023
Trino	Principal	hive-jdbc	Suppléments	3.1.3. amzn.11-1 . amzn2023
Zookeeper	Principal	emrfs	Suppléments	2.63.0-1 amzn2023
Zookeeper	Principal	hadoop-client	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-hdfs	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-hdfs-datanode	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-hdfs-journalnode	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-hdfs-namenode	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-httfs	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	Hadoop KMS	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-ma preduce	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-ma preduce-h istoryserver	Suppléments	3.3.6. amzn.4-1. amzn2023

Application	Type de demande	Package	Type de package	Version du package
Zookeeper	Principal	fil Hadoop	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-yarn-nodemanager	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-yarn-proxyserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-yarn-resourcemanager	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	hadoop-yarn-timelineserver	Suppléments	3.3.6. amzn.4-1. amzn2023
Zookeeper	Principal	ranger-km	Suppléments	2.0.0-1 mai 2023

Versions de l'application 7.2.0

Les applications suivantes sont prises en charge dans cette version : [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDKpour Java	2,23,18, 1,12,705	2,23,18, 1,12,656	2.20.160-amzn-0, 1.12.569
Python	3.9	3.9	3.9
Scala	2,1,17	2,1,17	2,1,17
AmazonClo udWatchAgent	1,300032,2	1,300032,2	1,300031,1
Delta	3.1.0	3.0.0	3.0.0
Flink	1.18.1	1.18.1	1.18.0
HBase	2.4,17	2.4,17	2.4,17
HCatalog	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6	3.3.6
Hive	3.1.3	3.1.3	3.1.3
Hudi	0,14.1-amzn-1	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0	4.11.0
Iceberg	1,5.0-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseG ateway	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1
MXNet	1.9.1	1.9.1	1.9.1
Oozie	5.2.1	5.2.1	5.2.1

	emr-7.2.0	emr-7.1.0	emr-7.0.0
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0
Presto	0,285	0,284	0,283
Spark	3.5.1	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0,1,2	0,1,2	0,1,2
Trino (PrestoSQL)	436	435	426
Zeppelin	0,1,1	0,1,1	0,1,1
ZooKeeper	3.9.1	3.9.1	3.5.10

notes de publication de la version 7.2.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 7.2.0 d'Amazon. Les modifications concernent la version 7.2.0.

Nouvelles fonctionnalités

- Mises à niveau des applications –Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0.
- Amazon EMR ajoute un support afin que vous puissiez utiliser d'autres applications telles que HBase Flink et Hive avec la classe de stockage Amazon S3 Express One Zone.
- Cette version ajoute la capacité de lire les objets restaurés, afin que vous puissiez lire les objets Glacier depuis un emplacement S3 avec le S3A protocole. Cette fonctionnalité fonctionne avec Spark, Flink et Hive.
- Étiquettes de nœuds avec dimensionnement géré : utilisez des étiquettes de nœuds avec vos instances en fonction du type de marché ou du type de nœud pour améliorer le dimensionnement

automatique d'AmazonEMR. Pour plus d'informations, consultez la section [Utilisation du dimensionnement géré sur Amazon EMR](#).

Problèmes connus

- Python 3.11 n'est pas pris en charge avec EMR Studio.

Modifications, améliorations et problèmes résolus

- Cette version corrige un problème de blocage qui peut survenir lors des opérations internes de nettoyage des étapes. Cette opération gère le cycle de vie des étapes au fur et à mesure qu'elles se terminent sur le EMR cluster. Ce problème affecte les EMR opérations critiques d'Amazon, telles que le fonctionnement des étapes et le dimensionnement.
- Cette version résout un problème AMIs selon lequel les clusters personnalisés dotés de certains fichiers journaux préexistants peuvent provoquer l'échec du démon de gestion des EMR journaux Amazon.
- Amazon EMR 7.2.0 met à niveau le EMR daemon Amazon responsable de la gestion des clusters et des activités de surveillance de la version AWS SDK 1 à la version 2.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge	
2023,52140708,0	6,196-102,177	mai 2023	8 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), Canada Ouest (Calgary), AWS GovCloud (États-Unis Ouest), AWS GovCloud (États-Unis Est), Chine (Pékin), Chine (Ningxia)

Versions Java 7.2.0 par défaut

Les EMR versions 7.2 et ultérieures d'Amazon sont fournies avec Amazon Corretto 17 (basé sur JDK Open) par défaut pour les applications compatibles avec Corretto 17 JDK (17), à l'exception d'Apache Livy.

Le tableau suivant indique les versions Java par défaut pour les applications dans Amazon EMR 7.2.0. Si vous souhaitez modifier la valeur par défaut JVM de votre cluster, suivez les instructions fournies [Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique](#) pour chaque application exécutée sur le cluster. Vous ne pouvez utiliser qu'une seule version d'exécution de Java pour un cluster. Amazon EMR ne prend pas en charge l'exécution de différents nœuds ou applications sur différentes versions d'exécution sur le même cluster.

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Delta	17, 11, 8
Flink	17, 11, 8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Étincelle RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versions des composants 7.2.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
adot-java-agent	1,31,0	Agent Java qui collecte des métriques à partir des démons d'application.

Composant	Version	Description
delta	3.1.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
emr-amazon-cloudwatch-agent	1,300032,2-amzn-0	Une application qui collecte des métriques internes au niveau du système et des métriques d'application personnalisées à partir d'instances AmazonEC2.
emr-ddb	5.4.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.11.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3,15,0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.16.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.32.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.11.0	EMRConnecteur S3Select
emr-wal-cli	1.2.2	CLI utilisée pour la liste/suppression EMR WAL.

Composant	Version	Description
emrfs	2,63,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1,18.1-amzn-1	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1,18.1-amzn-1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
hadoop-client	3.3.6-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.6-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.6-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.6-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.6-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	3.3.6-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.6-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	3.3.6-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.6-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.6-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.6-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2,4,17-amzn-6	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2,4,17-amzn-6	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2,4,17-amzn-6	HBaseclient en ligne de commande.
hbase-rest-server	2,4,17-amzn-6	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2,4,17-amzn-6	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hbase-operator-tools	2,4,17-amzn-6	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-11	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-11	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-11	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-11	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-11	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-11	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-11	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0,14.1-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.

Composant	Version	Description
hudi-presto	0,14.1-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0,14.1-amzn-1	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0,14.1-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1,5.0-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.5.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.8.0 - incubation	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.

Composant	Version	Description
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.7.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.3	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.3	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	6.0.0	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0,285-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0,285-amzn-0	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
presto-client	0,285-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	436-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	436-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	436-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.3.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.5.1-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.5.1-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	3.5.1-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.5.1-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	24.02.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-9	L'YARN application Tez et les bibliothèques.
tez-on-worker	0.10.2-amzn-9	L'YARN application tez et les bibliothèques pour les nœuds de travail.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.9.1-amzn-1	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.9.1-amzn-1	ZooKeeper client en ligne de commande.

Classifications de configuration 7.2.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour de plus amples informations, veuillez consulter [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

classifications EMR-7.2.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the Resource Manager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Classifications	Description	Actions de reconfiguration
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/ Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.

Classifications	Description	Actions de reconfiguration
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.

Classifications	Description	Actions de reconfiguration
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Modifier la page d'accueil Java de KMS Hadoop	Not available.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.

Classifications	Description	Actions de reconfiguration
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.

Classifications	Description	Actions de reconfiguration
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier <code>lakeformation.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-delta	Modification des valeurs dans le fichier <code>delta.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier <code>exchange-manager.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier <code>iceberg.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifiez les valeurs dans le fichier <code>hudi.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-logback	Modifiez les valeurs dans ZooKeeper le fichier logback.xml.	Restarts Zookeeper server.
emr-metrics	Modifiez les paramètres métriques EMR pour ce nœud.	Restarts the CloudWatchAgent service.

journal des modifications 7.2.0

Journal des modifications pour la version 7.2.0 et notes de version

Date	Événement	Description
25/07/2024	Publication de documents	Notes de mise à jour d'Amazon EMR 7.2.0 publiées pour la première fois
16/07/2024	Première version	Amazon EMR 7.2.0 a été déployé pour la première fois dans les régions commerciales initiales

Amazon EMR version 7.1.0

Versions de l'application 7.1.0

Les applications suivantes sont prises en charge dans cette version : [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#),

[JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDKpour Java	2,23,18, 1,12,705	2,23,18, 1,12,656	2.20.160-amzn-0, 1.12.569
Python	3.9	3.9	3.9
Scala	2,1,17	2,1,17	2,1,17
AmazonClo udWatchAgent	1,300032,2	1,300032,2	1,300031,1
Delta	3.1.0	3.0.0	3.0.0
Flink	1.18.1	1.18.1	1.18.0
HBase	2.4,17	2.4,17	2.4,17
HCatalog	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6	3.3.6
Hive	3.1.3	3.1.3	3.1.3

	emr-7.2.0	emr-7.1.0	emr-7.0.0
Hudi	0,14.1-amzn-1	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0	4.11.0
Iceberg	1,5.0-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1
MXNet	1.9.1	1.9.1	1.9.1
Mahout	-	-	-
Oozie	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0
Presto	0,285	0,284	0,283
Spark	3.5.1	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0,1,2	0,1,2	0,1,2
Trino (PrestoSQL)	436	435	426
Zeppelin	0,1,1	0,1,1	0,1,1
ZooKeeper	3.9.1	3.9.1	3.5.10

notes de publication de la version 7.1.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 7.1.0 d'Amazon. Les modifications concernent la version 7.1.0.

Nouvelles fonctionnalités

- Mises à niveau des applications –Amazon EMR 7.1.0 application upgrades include Livy 0.8.0, Trino 435, and ZooKeeper 3.9.1.
- [Remplacement de nœuds](#) défectueux — Avec Amazon EMR 7.1.0 et versions ultérieures, le remplacement de nœuds défectueux est activé par défaut. Amazon EMR remplacera donc facilement vos nœuds défectueux. Pour éviter d'affecter vos flux de travail existants sur les EMR versions 7.0.0 et antérieures d'Amazon, le remplacement des nœuds défectueux est désactivé si vous avez activé la protection contre la résiliation dans votre cluster.
- [CloudWatch Agent](#) : configurez l' CloudWatch agent pour utiliser des métriques système supplémentaires, ajouter des métriques d'application et modifier la destination des métriques avec la EMR configuration AmazonAPI.

Problèmes connus

- Python 3.11 n'est pas pris en charge avec EMR Studio.

Modifications, améliorations et problèmes résolus

- Alors qu'Amazon EMR 7.1.0 prend en charge Python 3.9 par défaut, Livy 0.8.0 et Spark d'Amazon EMR 7.1.0 prennent en charge Python 3.11.
- Cette version corrige le problème de devoir exécuter chaque ligne une par une lors de l'utilisation PySpark avec Python version 3.11.
- Mise à niveau de Zeppelin — Amazon EMR 7.1.0 inclut une mise à niveau de Zeppelin vers la version pour Java v2. AWS SDK Cette mise à niveau permet à un ordinateur portable Zeppelin S3 d'accepter un fournisseur de matériel de chiffrement personnalisé. Le AWS SDK for Java v2 supprime l' EncryptionMaterialsProvider interface. Lorsque vous passez à Amazon EMR 7.1.0, vous devez implémenter l'[interface Keyring](#) si vous souhaitez utiliser un chiffrement personnalisé. Pour un exemple de mise en œuvre de l'interface Keyring, consultez [KmsKeyring.java](#).
- Lors de la mise à niveau vers la EMR version 7.1.0 d'Amazon, modifiez votre fournisseur de clés personnalisé pour le chiffrement du disque local afin de générer des clés à l'aide de l'AESalgorithmeAES/GCM/NoPadding. Si vous ne mettez pas à jour l'algorithme, la création du

cluster risque d'échouer avec le message d'erreur. `Local disk encryption failed on master instance (i-123456789) due to internal error.` Pour plus d'informations sur la création d'un fournisseur de clés personnalisé, voir [Création d'un fournisseur de clés personnalisé](#).

- Amazon EMR 7.1.0 améliore la résilience d'un nœud dans des conditions d'espace disque restreint en améliorant la logique de troncature des journaux pour les fichiers dont les descripteurs de fichier sont ouverts.
- Cette version améliore la logique de codage et de décodage afin de minimiser le risque de corruption des données et de défaillance du nœud, les EMR démons Amazon lisant et écrivant des fichiers lors du redémarrage d'un nœud.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023,520-40708,0	6,196-102,177 mai 2023	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Pacifique) (Sydney), Asie-Pacifique (Jakarta) , Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023.3.240219,0	6.1.77-99,164 mai 2023	8 mai 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), Canada Ouest (Calgary), AWS GovCloud (États-Unis)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Ouest), AWS GovCloud (États-Unis Est), Chine (Pékin), Chine (Ningxia)

Versions Java 7.1.0 par défaut

Les EMR versions 7.1 et ultérieures d'Amazon sont fournies avec Amazon Corretto 17 (basé sur JDK Open) par défaut pour les applications compatibles avec Corretto 17 JDK (17), à l'exception d'Apache Livy.

Le tableau suivant indique les versions Java par défaut pour les applications dans Amazon EMR 7.1.0. Si vous souhaitez modifier la valeur par défaut JVM de votre cluster, suivez les instructions fournies [Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique](#) pour chaque application exécutée sur le cluster. Vous ne pouvez utiliser qu'une seule version d'exécution de Java pour un cluster. Amazon EMR ne prend pas en charge l'exécution de différents nœuds ou applications sur différentes versions d'exécution sur le même cluster.

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Delta	17, 11, 8
Flink	17, 11, 8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Étincelle RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versions des composants 7.1.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
adot-java-agent	1,31,0	Agent Java qui collecte des métriques à partir des démons d'application.
delta	3.0.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
emr-amazon-cloudwatch-agent	1,300032,2-amzn-0	Une application qui collecte des métriques internes au niveau du système et des métriques d'application personnalisées à partir d'instances AmazonEC2.
emr-ddb	5.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.10.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3,14,0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.31.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.10.0	EMRConnecteur S3Select
emr-wal-cli	1.2.2	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,62,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1,18.1-amzn-0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1,18.1-amzn-0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
hadoop-client	3.3.6-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.6-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.6-amzn-3	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	3.3.6-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.6-amzn-3	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	3.3.6-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.6-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.6-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.6-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.6-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.6-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	2,4,17-amzn-5	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2,4,17-amzn-5	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2,4,17-amzn-5	HBaseclient en ligne de commande.
hbase-rest-server	2,4,17-amzn-5	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2,4,17-amzn-5	Service fournissant un point de terminaison Thrift pour. HBase
hbase-operator-tools	2,4,17-amzn-5	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-10	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-10	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-10	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-10	Client de ligne de commande Hive.

Composant	Version	Description
hive-hbase	3.1.3-amzn-10	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-10	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-10	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.14.1-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.14.1-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.14.1-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.14.1-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.4.3-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.

Composant	Version	Description
jupyterhub	1.5.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.8.0 - incubation	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.7.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.3	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.3	Connecteurs Apache Phoenix pour Spark-3

Composant	Version	Description
phoenix-query-server	5.1.3	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0,284-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0,284-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0,284-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	435-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	435-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	435-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.

Composant	Version	Description
r	4.3.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.5.0-amzn-1	Clients de ligne de commande Spark.
spark-history-server	3.5.0-amzn-1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.5.0-amzn-1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.5.0-amzn-1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	23.10.0-amzn-1	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-8	L'YARNapplication Tez et les bibliothèques.

Composant	Version	Description
tez-on-worker	0.10.2-amzn-8	L'YARN application tez et les bibliothèques pour les nœuds de travail.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.9.1-amzn-0	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.9.1-amzn-0	ZooKeeper client en ligne de commande.

Classifications de configuration 7.1.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

classifications EMR-7.1.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.

Classifications	Description
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
docker-conf	Modifiez les paramètres liés au docker.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog <code>proto-hive-site</code> fichier <code>.xml</code> .
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat <code>log4j2.properties</code> de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier <code>webhcat-site.xml</code> de HCatalog W.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-java-home	Modifier la page d'accueil Java de KMS Hadoop
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier lakeformation.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.

Classifications	Description
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.

Classifications	Description
trino-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Trino.
trino-connector-delta	Modification des valeurs dans le fichier delta.properties de Trino.
trino-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Trino.
trino-exchange-manager	Modification des valeurs dans le fichier exchange-manager.properties de Trino.
trino-connector-iceberg	Modifiez les valeurs dans le fichier iceberg.properties de Trino.
trino-connector-hudi	Modifiez les valeurs dans le fichier hudi.properties de Trino.
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.

Classifications	Description
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.

Classifications	Description
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-logback	Modifiez les valeurs dans ZooKeeper le fichier logback.xml.

Classifications	Description
emr-metrics	Modifiez les paramètres métriques EMR pour ce nœud.

journal des modifications 7.1.0

Journal des modifications pour la version 7.1.0 et notes de version

Date	Événement	Description
2024-05-10	Publication de documents	Notes de mise à jour d'Amazon EMR 7.1.0 publiées pour la première fois
08-05-2024/	Déploiement terminé	Amazon EMR 7.1.0 entièrement déployé dans toutes les régions prises en charge
23/04/2024-04	Première version	Amazon EMR 7.1.0 a été déployé pour la première fois dans les régions commerciales initiales

Amazon EMR version 7.0.0

Versions d'application 7.0.0

Les applications suivantes sont prises en charge dans cette version : [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDKpour Java	2,23,18, 1,12,705	2,23,18, 1,12,656	2.20.160-amzn-0, 1.12.569
Python	3.9	3.9	3.9
Scala	2,1,17	2,1,17	2,1,17
AmazonClo udWatchAgent	1,300032,2	1,300032,2	1,300031,1
Delta	3.1.0	3.0.0	3.0.0
Flink	1.18.1	1.18.1	1.18.0
HBase	2.4,17	2.4,17	2.4,17
HCatalog	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.6	3.3.6
Hive	3.1.3	3.1.3	3.1.3
Hudi	0,14.1-amzn-1	0.14.1-amzn-0	0.14.0-amzn-1
Hue	4.11.0	4.11.0	4.11.0
Iceberg	1,5.0-amzn-0	1.4.3-amzn-0	1.4.2-amzn-0
JupyterEnterpriseG ateway	2.6.0	2.6.0	2.6.0

	emr-7.2.0	emr-7.1.0	emr-7.0.0
JupyterHub	1.5.0	1.5.0	1.5.0
Livy	0.8.0	0.8.0	0.7.1
MXNet	1.9.1	1.9.1	1.9.1
Mahout	-	-	-
Oozie	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0
Presto	0,285	0,284	0,283
Spark	3.5.1	3.5.0	3.5.0
Sqoop	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0,1,2	0,1,2	0,1,2
Trino (PrestoSQL)	436	435	426
Zeppelin	0,1,1	0,1,1	0,1,1
ZooKeeper	3.9.1	3.9.1	3.5.10

Notes de mise à jour 7.0.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 7.0.0 d'Amazon. Les modifications ont été apportées à la version 6.15.0.

Nouvelles fonctionnalités

- Mises à niveau des applications –Amazon EMR 7.0.0 application upgrades include Python 3.9, Spark 3.5, Flink 1.18, and Delta 3.0. Cette version ajoute également la prise en charge de l'application [CloudWatch Agent Amazon](#) et abandonne la prise en charge de Ganglia.
- Amazon Corretto 17 — Les versions 7.0 et ultérieures d'EMR Amazon sont fournies avec Amazon Corretto 17 (basé sur JDK Open) par défaut pour les applications compatibles avec Corretto 17 (17), à l'exception d'Apache Livy. JDK Pour plus d'informations sur les JDK versions prises en charge pour les applications de cette version, consultez [Versions Java par défaut 7.0.0](#).
- Amazon Linux 2023 — Avec la version 7.0, les EMR clusters Amazon s'exécutent désormais sur AL2 023 par défaut. Pour plus d'informations sur la manière dont cela affecte la AMI version par défaut, consultez la section [Considérations relatives aux mises à jour logicielles](#) dans le guide EMR de gestion Amazon. Notez également que AL2 023 a supprimé Python 2.7, donc tous les composants nécessitant Python doivent désormais être écrits avec Python 3.
- S3 sur Outposts avec — s3a Amazon prend EMR désormais en charge les compartiments Amazon S3 on Outposts avec le système de fichiers. s3a Pour plus d'informations sur S3 sur Outposts, voir la rubrique [Qu'est-ce que S3 sur Outposts ?](#) du Guide de l'utilisateur Amazon Simple Storage Service.

Problèmes connus

- Notez que vous ne pouvez pas exécuter plus d'une ligne à la fois si vous utilisez PySpark Python 3.10 ou version ultérieure. Vous devez exécuter chaque ligne une par une.

Modifications, améliorations et problèmes résolus

- Pour maintenir l'état de toutes les instances d'un cluster, Amazon EMR intègre Apache YARN, Apache HDFS et Kerberos. Avec la version 7.0, nous avons amélioré ces intégrations afin de supprimer en toute sécurité l'état des instances arrêtées en raison d'une mise à l'échelle ou d'autres opérations. Cette amélioration est particulièrement importante pour les clusters longue durée pour lesquels la mise à l'échelle gérée est activée, car font régulièrement l'objet de mises à l'échelle et accumulent des milliers d'instances arrêtées au cours de leur durée de vie.
- Cette version améliore la configuration de Kerberos afin d'inclure uniquement la prise en charge des chiffrements AES basés. Les Kerberos KDC utilisant des chiffrements non AES basés ne sont plus pris en charge avec les EMR clusters exécutés sur Amazon EMR versions 7.0.0 et

supérieures. AESLes chiffrements basés sur la technologie offrent la sécurité la plus élevée pour vos clusters.

- Dans le cadre de la migration vers la version AWS SDK 2.x, Amazon EMR 7.0 inclut une mise à jour du connecteur Spark Kinesis pour des raisons de compatibilité. Cette mise à jour n'est pas disponible dans la version communautaire d'Apache Spark. Si vous utilisez le connecteur Spark Kinesis à partir d'une EMR version Amazon inférieure à 7.0, vous devez migrer les codes de vos applications pour qu'ils s'exécutent sur SDK 2.x avant de pouvoir migrer vos charges de travail vers Amazon 7.0. EMR Pour de plus amples informations, veuillez consulter [Migration du connecteur Spark Kinesis vers la version 2.x SDK pour Amazon 7.0 EMR](#).
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023,52140708,0	6,196-102,177 mai 2023	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), Canada Ouest (Calgary), AWS GovCloud (États-Unis Ouest), AWS GovCloud (États-Unis Est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023.3.240304,0	6.1.79-99,164 mai 2023	12 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), Canada Ouest (Calgary), AWS GovCloud (États-Unis)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Ouest), AWS GovCloud (États-Unis Est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023.3.2_240219,1	6.1.77-99,164 mai 2023	1er mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), Canada Ouest (Calgary), AWS GovCloud (États-Unis)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Ouest), AWS GovCloud (États-Unis Est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023.3.240205,0	6,175-99,163 mai 2023	19 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), Canada Ouest (Calgary), AWS GovCloud (États-Unis)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Ouest), AWS GovCloud (États-Unis Est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023.3.2_240122.l	6,172-96,166 mai 2023	5 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), Canada Ouest (Calgary), AWS GovCloud (États-Unis)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Ouest), AWS GovCloud (États-Unis Est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023.3.240108.0	6,172-96,166 mai 2023	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), Canada Ouest (Calgary), AWS GovCloud (États-Unis)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Ouest), AWS GovCloud (États-Unis Est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023.3.2 23 1211,4	6.1.66-91.160.amzn2023	19 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

Versions Java par défaut 7.0.0

Les EMR versions 7.0 et ultérieures d'Amazon sont fournies avec Amazon Corretto 17 (basé sur JDK Open) par défaut pour les applications compatibles avec Corretto 17 JDK (17), à l'exception d'Apache Livy.

Le tableau suivant indique les versions Java par défaut pour les applications dans Amazon EMR 7.0.0. Si vous souhaitez modifier la valeur par défaut JVM de votre cluster, suivez les instructions fournies [Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique](#) pour chaque application exécutée sur le cluster. Vous ne pouvez utiliser qu'une seule version d'exécution de Java pour un cluster. Amazon EMR ne prend pas en charge l'exécution de différents nœuds ou applications sur différentes versions d'exécution sur le même cluster.

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Delta	17, 11, 8
Flink	11, 8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Étincelle RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versions des composants 7.0.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion-amzn-EmrVersion.EmrVersion*

commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
adot-java-agent	1,31,0	Agent Java qui collecte des métriques à partir des démons d'application.
delta	3.0.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
delta-standalone-connectors	0.6.0	Les connecteurs Delta fournissent différents environnements d'exécution pour intégrer Delta Lake avec des moteurs tels que Flink, Hive et Presto.
emr-amazon-cloudwatch-agent	1.300031.1-amzn-0	Une application qui collecte des métriques internes au niveau du système et des métriques d'application personnalisées à partir d'instances AmazonEC2.
emr-ddb	5.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.9.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3,13,0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.30.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.9.0	EMRConnecteur S3Select
emr-wal-cli	1.2.1	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,61,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.18.0-amzn-0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.18.0-amzn-0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
hadoop-client	3.3.6-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.6-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	3.3.6-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.6-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.6-amzn-2	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.3.6-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.6-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.6-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.6-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.6-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.6-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	2.4.17-amzn-4	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.17-amzn-4	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.17-amzn-4	HBaseclient en ligne de commande.
hbase-rest-server	2.4.17-amzn-4	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.17-amzn-4	Service fournissant un point de terminaison Thrift pour HBase
hbase-operator-tools	2.4.17-amzn-4	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-9	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-9	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-9	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-9	Client de ligne de commande Hive.

Composant	Version	Description
hive-hbase	3.1.3-amzn-9	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-9	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-9	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.14.0-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.14.0-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.14.0-amzn-1	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.14.0-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.4.2-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.

Composant	Version	Description
jupyterhub	1.5.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.7.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.3	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.3	Connecteurs Apache Phoenix pour Spark-3

Composant	Version	Description
phoenix-query-server	5.1.3	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.283-amzn-1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.283-amzn-1	Service pour exécuter les éléments d'une requête.
presto-client	0.283-amzn-1	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	426-amzn-1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	426-amzn-1	Service pour exécuter les éléments d'une requête.
trino-client	426-amzn-1	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.

Composant	Version	Description
r	4.1.3	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.5.0-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.5.0-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.5.0-amzn-0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.5.0-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	23.10.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-7	L'YARNapplication Tez et les bibliothèques.

Composant	Version	Description
tez-on-worker	0.10.2-amzn-7	L'YARN application tez et les bibliothèques pour les nœuds de travail.
webserver	2,4,58	HTTP serveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 7.0.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

Classifications emr-7.0.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.

Classifications	Description
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
docker-conf	Modifiez les paramètres liés au docker.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.

Classifications	Description
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.

Classifications	Description
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-java-home	Modifier la page d'accueil Java de KMS Hadoop
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.

Classifications	Description
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier lakeformation.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.

Classifications	Description
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.

Classifications	Description
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.
trino-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Trino.
trino-connector-delta	Modification des valeurs dans le fichier delta.properties de Trino.
trino-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Trino.
trino-exchange-manager	Modification des valeurs dans le fichier exchange-manager.properties de Trino.
trino-connector-iceberg	Modifiez les valeurs dans le fichier iceberg.properties de Trino.
trino-connector-hudi	Modifiez les valeurs dans le fichier hudi.properties de Trino.
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.

Classifications	Description
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.

Classifications	Description
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Journal des modifications 7.0.0

Journal des modifications pour la version 7.0.0 et les notes de mise à jour

Date	Événement	Description
29/12/2023	Déploiement terminé	Amazon EMR 7.0.0 entièrement déployé dans toutes les régions prises en charge
28/12/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 7.0.0 publiées pour la première fois
19/12/2023	Première version	Amazon EMR 7.0.0 a été déployé pour la première fois dans les régions commerciales initiales

Versions publiées par Amazon EMR 6.x

Cette section contient les versions des applications, les notes de mise à jour, les versions des composants et les classifications de configuration disponibles dans chaque version d'Amazon EMR 6.x.

Lorsque vous lancez un cluster, vous pouvez choisir parmi plusieurs versions d'AmazonEMR. Vous pouvez ainsi tester et utiliser des versions d'applications qui répondent à vos exigences de compatibilité. Vous spécifiez le numéro de version avec l'étiquette de version. Les étiquettes de version sont sous la forme `emr-x.x.x`. Par exemple, `emr-7.2.0`.

Les nouvelles EMR versions d'Amazon sont mises à disposition dans différentes régions sur une période de plusieurs jours, en commençant par la première région à la date de sortie initiale. Il est possible que la dernière version ne soit pas disponible dans votre région pendant cette période.

Pour un tableau complet des versions des applications dans chaque version d'Amazon EMR 6.x, consultez [Versions des applications dans les versions 6.x d'Amazon EMR](#).

Rubriques

- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Amazon EMR version 6.15.0](#)
- [Amazon EMR version 6.14.0](#)
- [Amazon EMR version 6.13.0](#)
- [Amazon EMR version 6.12.0](#)
- [Amazon EMR version 6.11.1](#)
- [Amazon EMR version 6.11.0](#)
- [Amazon EMR version 6.10.1](#)
- [Amazon EMR version 6.10.0](#)
- [Amazon EMR version 6.9.1](#)
- [Amazon EMR version 6.9.0](#)
- [Amazon EMR version 6.8.1](#)
- [Amazon EMR version 6.8.0](#)
- [Amazon EMR version 6.7.0](#)
- [Amazon EMR version 6.6.0](#)
- [Amazon EMR version 6.5.0](#)
- [Amazon EMR version 6.4.0](#)
- [Amazon EMR version 6.3.1](#)
- [Amazon EMR version 6.3.0](#)
- [Amazon EMR version 6.2.1](#)
- [Amazon EMR version 6.2.0](#)
- [Amazon EMR version 6.1.1](#)
- [Amazon EMR version 6.1.0](#)
- [Amazon EMR version 6.0.1](#)
- [Amazon EMR version 6.0.0](#)

Versions des applications dans les versions 6.x d'Amazon EMR

Pour obtenir un tableau complet répertoriant les versions d'applications disponibles dans chaque version d'Amazon EMR 6.x, ouvrez [Versions des applications dans les versions 6.x d'Amazon EMR](#) dans votre navigateur.

Amazon EMR version 6.15.0

Versions d'application 6.15.0

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Kit AWS SDK pour Java	2.20.160-amzn-0, 1.12.569	1,12,543	1,12,513	1,12,490
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,17	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Delta	2.4.0	2.4.0	2.4.0	2.4.0
Flink	1,17.1	1,17.1	1.17.0	1.17.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4,17	2.4,17	2.4,17	2.4,17
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.6	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.14.0-amzn-0	0.13.1-amzn-2	0.13.1-amzn-1	0.13.1-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.4.0-amzn-0	1.3.1-amzn-0	1.3.0-amzn-1	1.3.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.5.0	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3	5.1.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,283	0,281	0,281	0,281
Spark	3.4.1	3.4.1	3.4.1	3.4.0

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,1,2	0,1,2	0,1,2	0,1,2
Trino (PrestoSQL)	426	422	414	414
Zeppelin	0,1,1	0,1,1	0,1,1	0,1,1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notes de mise à jour 6.15.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.15.0 d'Amazon. Les modifications ont été apportées à la version 6.14.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.15.0](#).

Nouvelles fonctionnalités

- Mises à niveau des applications –Amazon EMR 6.15.0 application upgrades include Apache Hadoop 3.3.6, Apache Hudi 0.14.0-amzn-0, Iceberg 1.4.0-amzn-0, and Trino 426..
- Des [lancements plus rapides pour les EMR clusters qui s'exécutent sur EC2](#) : le lancement d'un EC2 cluster Amazon EMR on est désormais jusqu'à 35 % plus rapide. Grâce à cette amélioration, la plupart des clients peuvent lancer leurs clusters en moins de 5 minutes.
- [CodeWhisperer pour EMR Studio](#) — Vous pouvez désormais utiliser Amazon CodeWhisperer avec Amazon EMR Studio pour obtenir des recommandations en temps réel lorsque vous écrivez du code JupyterLab. CodeWhisperer peut compléter vos commentaires, terminer des lignes de code individuelles, faire line-by-line des recommandations et générer des fonctions entièrement formées.
- Temps de [redémarrage des tâches plus rapides avec Flink](#) — Avec Amazon EMR 6.15.0 et versions ultérieures, plusieurs nouveaux mécanismes sont disponibles pour Apache Flink afin d'améliorer le temps de redémarrage des tâches lors des opérations de restauration ou de dimensionnement des tâches. Cela permet d'optimiser la vitesse de récupération et de redémarrage des graphes d'exécution afin d'améliorer la stabilité des tâches.

- [Contrôle d'accès détaillé et au niveau des tables pour les formats de tables ouvertes : avec Amazon EMR 6.15.0 et versions ultérieures, lorsque vous exécutez des tâches Spark sur Amazon EMR sur des clusters EC2 qui accèdent aux données du catalogue de données AWS Glue, vous pouvez les utiliser AWS Lake Formation pour appliquer des autorisations au niveau des tables, des lignes, des colonnes et des cellules aux tables basées](#) sur Hudi, Iceberg ou Delta Lake.
- Mise à niveau vers Hadoop — Amazon EMR 6.15.0 inclut une mise à niveau d'Apache Hadoop vers la version 3.3.6. Hadoop 3.3.6 était la dernière version au moment du déploiement d'Amazon EMR 6.15, publiée par Apache en juin 2023. Les versions précédentes d'Amazon EMR (6.9.0 à 6.14.x) utilisaient Hadoop 3.3.3.

La mise à niveau inclut des centaines d'améliorations et de correctifs, ainsi que des fonctionnalités telles que des paramètres de nœud de données reconfigurables, la DFSAdmin possibilité de lancer des opérations de reconfiguration en masse sur tous les nœuds de données actifs et un vectorisé API qui permet aux lecteurs gourmands en recherche de spécifier plusieurs plages à lire. Hadoop 3.3.6 ajoute également le support HDFS APIs et la sémantique de son log write-ahead (WAL), afin qu'HBase puisse fonctionner sur d'autres implémentations de systèmes de stockage. Pour plus d'informations, consultez les journaux des modifications pour les versions [3.3.4](#), [3.3.5](#) et [3.3.6](#) dans la documentation Apache Hadoop.

- Support AWS SDK pour Java, version 2 - Les applications Amazon EMR 6.15.0 peuvent utiliser AWS SDK les versions Java [1.12.569](#) ou [2.20.160](#) si l'application prend en charge la version v2. La version 2.x AWS SDK pour Java est une réécriture majeure de la base de code de la version 1.x. Il repose sur Java 8+ et ajoute plusieurs fonctionnalités fréquemment demandées. Il s'agit notamment de la prise en charge des E/S non bloquantes et de la possibilité de connecter une HTTP implémentation différente lors de l'exécution. Pour plus d'informations, y compris un guide SDK de migration de Java v1 à v2, consultez le guide [AWS SDK pour Java version 2](#).

Modifications, améliorations et problèmes résolus

- Pour améliorer vos EMR clusters à haute disponibilité, cette version permet la connectivité aux EMR daemons Amazon sur des hôtes locaux qui utilisent des points de terminaison. IPv6
- Cette version permet à la version TLS 1.2 de communiquer avec les nœuds ZooKeeper provisionnés sur tous les nœuds principaux de votre cluster de haute disponibilité.
- Cette version améliore la gestion des fichiers journaux de ZooKeeper transactions conservés sur les nœuds principaux afin de minimiser les scénarios dans lesquels les fichiers journaux dépassent les limites et interrompent les opérations du cluster.

- Cette version rend la communication intra-nœud plus résiliente pour les clusters à haute disponibilité EMR. Cette amélioration réduit le risque d'échec des actions d'amorçage ou de démarrage du cluster.
- Dans Amazon EMR 6.15.0, Tez introduit des configurations que vous pouvez spécifier pour ouvrir de manière asynchrone les divisions d'entrée dans une division groupée Tez. Cela permet d'accélérer les performances des requêtes de lecture lorsqu'il existe un grand nombre de divisions d'entrée dans une seule division groupée Tez. Pour plus d'informations, consultez la rubrique [Ouverture fractionnée asynchrone Tez](#).
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap),

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	13 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

Versions Java par défaut 6.15.0

EMR Les versions 6.12.0 et ultérieures d'Amazon prennent en charge toutes les applications dotées d'Amazon Corretto 8 par défaut, à l'exception de Trino. Pour Trino, Amazon EMR prend en charge Amazon Corretto 17 par défaut à partir de la version 6.9.0 d'Amazon. EMR Amazon prend également en charge certaines applications avec Amazon Corretto 11 et 17. Ces applications sont répertoriées dans le tableau suivant. Si vous souhaitez modifier la valeur par défaut JVM de votre cluster, suivez les instructions fournies [Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique](#) pour chaque application exécutée sur le cluster. Vous ne pouvez utiliser qu'une seule version d'exécution de Java pour un cluster. Amazon EMR ne prend pas en charge l'exécution de différents nœuds ou applications sur différentes versions d'exécution sur le même cluster.

Bien qu'Amazon EMR prenne en charge Amazon Corretto 11 et 17 sur Apache Spark, Apache Hadoop et Apache Hive, les performances peuvent régresser pour certaines charges de travail lorsque vous utilisez ces versions de Corretto. Nous vous recommandons de tester vos charges de travail avant de modifier les valeurs par défaut.

Le tableau suivant indique les versions Java par défaut pour les applications dans Amazon EMR 6.15.0 :

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Étincelle RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versions des composants 6.15.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.4.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
<code>delta-standalone-connectors</code>	0.6.0	Les connecteurs Delta fournissent différents environnements d'exécution pour intégrer Delta Lake avec des moteurs tels que Flink, Hive et Presto.
<code>emr-ddb</code>	5.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-goodies	3.8.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.12.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.29.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.8.0	EMRConnecteur S3Select
emr-wal-cli	1.2.0	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,60,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.17.1-amzn-1	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.17.1-amzn-1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.3.6-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.6-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.6-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.6-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.6-amzn-1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	3.3.6-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.6-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	3.3.6-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.6-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.6-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.6-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.17-amzn-3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.17-amzn-3	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.17-amzn-3	HBaseclient en ligne de commande.
hbase-rest-server	2.4.17-amzn-3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.17-amzn-3	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hbase-operator-tools	2.4.17-amzn-3	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-8	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-8	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-8	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-8	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-8	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-8	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-8	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.14.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.

Composant	Version	Description
hudi-presto	0.14.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.14.0-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.14.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.4.0-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.5.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.

Composant	Version	Description
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.7.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.3	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.3	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.3	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.283-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.283-amzn-0	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
presto-client	0.283-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	426-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	426-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	426-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.4.1-amzn-2	Clients de ligne de commande Spark.
spark-history-server	3.4.1-amzn-2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	3.4.1-amzn-2	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.4.1-amzn-2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	23.08.1-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-6	L'YARNapplication Tez et les bibliothèques.
tez-on-worker	0.10.2-amzn-6	L'YARNapplication tez et les bibliothèques pour les nœuds de travail.
webserver	2.4.41+	HTTPserveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.15.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.15.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the Resource Manager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Actions de reconfiguration
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-PROTO-HIVE-SITE	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.

Classifications	Description	Actions de reconfiguration
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Modifier la page d'accueil Java de KMS Hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.

Classifications	Description	Actions de reconfiguration
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.

Classifications	Description	Actions de reconfiguration
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier lakeformation.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)

Classifications	Description	Actions de reconfiguration
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Trino.	Not available.
trino-connector-delta	Modification des valeurs dans le fichier delta.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier exchange-manager.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier iceberg.properties de Trino.	Restarts Trino-Server (for Trino)

Classifications	Description	Actions de reconfiguration
trino-connector-hudi	Modifiez les valeurs dans le fichier hudi.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.

Classifications	Description	Actions de reconfiguration
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.

Classifications	Description	Actions de reconfiguration
yarn-env	Modifiez les valeurs dans l'YARN environnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.15.0

Journal des modifications pour la version 6.15.0 et les notes de mise à jour

Date	Événement	Description
17/11/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 6.15.0 publiées pour la première fois
17/11/2023	Déploiement terminé	Amazon EMR 6.15.0 entièrement déployé dans toutes les régions prises en charge
13/11/2023	Première version	Amazon EMR 6.15.0 a été déployé pour la première fois dans les régions commerciales initiales

Amazon EMR version 6.14.0

Versions d'application 6.14.0

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
Kit AWS SDK pour Java	1,12,543	1,12,513	1,12,490	1,12,446
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.4.0	2.2.0
Flink	1,17.1	1.17.0	1.17.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4,17	2.4,17	2.4,17	2.4.15
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-2	0.13.1-amzn-1	0.13.1-amzn-0	0.13.0-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.3.1-amzn-0	1.3.0-amzn-1	1.3.0-amzn-0	1.2.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.3	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,281	0,281	0,281	0,279
Spark	3.4.1	3.4.1	3.4.0	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,1,2	0,1,2	0,1,2	0,1,2
Trino (PrestoSQL)	422	414	414	410
Zeppelin	0,1,1	0,1,1	0,1,1	0,1,1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notes de mise à jour 6.14.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.14.0 d'Amazon. Les modifications ont été apportées à la version 6.13.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.14.0](#).

Nouvelles fonctionnalités

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- [Amazon EMR Managed Scaling](#) est désormais disponible dans la région ap-southeast-3 Asie-Pacifique (Jakarta) pour les clusters que vous créez avec Amazon EMR 6.14.0 ou version ultérieure.

Modifications, améliorations et problèmes résolus

- La version 6.14.0 optimise la gestion des journaux avec Amazon exécuté EMR sur Amazon. EC2 C'est pourquoi vous constaterez peut-être une légère réduction des coûts de stockage pour les journaux de votre cluster.
- La version 6.14.0 améliore le flux de travail de dimensionnement afin de prendre en compte les différentes instances principales dont la taille varie considérablement par rapport à leurs volumes AmazonEBS. Cette amélioration s'applique uniquement aux nœuds principaux ; les opérations de réduction des nœuds de tâches ne sont pas affectées.
- La version 6.14.0 améliore la façon dont Amazon EMR interagit avec les applications open source telles que Apache Hadoop YARN ResourceManager and HDFS NameNode Cette amélioration réduit le risque de retards opérationnels liés à la mise à l'échelle du cluster et atténue les échecs de démarrage dus à des problèmes de connectivité avec les applications open source.
- La version 6.14.0 optimise l'installation des applications au lancement du cluster. Cela améliore les temps de démarrage du cluster pour certaines combinaisons d'EMR applications Amazon.
- La version 6.14.0 résout un problème selon lequel les opérations de réduction du cluster pouvaient se bloquer lorsqu'un cluster exécuté dans un VPC domaine personnalisé redémarre un nœud principal ou un nœud de tâche.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfor

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			<p>t), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)</p>

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	17 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)
2,0.2023 906.0	4,1,4322	11 septembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

Versions Java par défaut 6.14.0

EMR Les versions 6.12.0 et ultérieures d'Amazon prennent en charge toutes les applications dotées d'Amazon Corretto 8 par défaut, à l'exception de Trino. Pour Trino, Amazon EMR prend en charge Amazon Corretto 17 par défaut à partir de la version 6.9.0 d'Amazon. EMR Amazon prend également en charge certaines applications avec Amazon Corretto 11 et 17. Ces applications sont répertoriées dans le tableau suivant. Si vous souhaitez modifier la valeur par défaut JVM de votre cluster, suivez les instructions fournies [Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique](#) pour chaque application exécutée sur le cluster. Vous ne pouvez utiliser qu'une seule version d'exécution de Java pour un cluster. Amazon EMR ne prend pas en charge l'exécution de différents nœuds ou applications sur différentes versions d'exécution sur le même cluster.

Bien qu'Amazon EMR prenne en charge Amazon Corretto 11 et 17 sur Apache Spark, Apache Hadoop et Apache Hive, les performances peuvent régresser pour certaines charges de travail lorsque vous utilisez ces versions de Corretto. Nous vous recommandons de tester vos charges de travail avant de modifier les valeurs par défaut.

Le tableau suivant indique les versions Java par défaut pour les applications dans Amazon EMR 6.14.0 :

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Étincelle RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versions des composants 6.14.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion`

commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.4.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
delta-standalone-connectors	0.6.0	Les connecteurs Delta fournissent différents environnements d'exécution pour intégrer Delta Lake avec des moteurs tels que Flink, Hive et Presto.
emr-ddb	5.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.7.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.11.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise

Composant	Version	Description
emr-s3-dist-cp	2.28.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.7.0	EMRConnecteur S3Select
emr-wal-cli	1.1.0	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,59,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.17.1-amzn-0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.17.1-amzn-0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	3.3.3-amzn-6	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.3-amzn-6	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.3-amzn-6	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.3-amzn-6	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.3-amzn-6	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httfs-server	3.3.3-amzn-6	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.3-amzn-6	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-6	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-6	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	3.3.3-amzn-6	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.3-amzn-6	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.17-amzn-2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.17-amzn-2	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.17-amzn-2	HBaseclient en ligne de commande.
hbase-rest-server	2.4.17-amzn-2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.17-amzn-2	Service fournissant un point de terminaison Thrift pour HBase
hbase-operator-tools	2.4.17-amzn-2	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-7	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	3.1.3-amzn-7	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-7	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-7	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-7	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-7	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-7	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.13.1-amzn-2	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.13.1-amzn-2	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.13.1-amzn-2	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.

Composant	Version	Description
hudi-spark	0.13.1-amzn-2	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.3.1-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.5.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.7.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.3	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.3	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.3	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.281-amzn-2	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.281-amzn-2	Service pour exécuter les éléments d'une requête.
presto-client	0.281-amzn-2	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	422-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.

Composant	Version	Description
trino-worker	422-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	422-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.4.1-amzn-1	Clients de ligne de commande Spark.
spark-history-server	3.4.1-amzn-1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.4.1-amzn-1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.4.1-amzn-1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	23.06.0-amzn-2	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.

Composant	Version	Description
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-5	L'YARN application Tez et les bibliothèques.
tez-on-worker	0.10.2-amzn-5	L'YARN application tez et les bibliothèques pour les nœuds de travail.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.14.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.14.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Modifier la page d'accueil Java de KMS Hadoop	Not available.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.

Classifications	Description	Actions de reconfiguration
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.

Classifications	Description	Actions de reconfiguration
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier <code>lakeformation.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-delta	Modification des valeurs dans le fichier <code>delta.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier <code>exchange-manager.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier <code>iceberg.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifiez les valeurs dans le fichier <code>hudi.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.14.0

Journal des modifications pour la version 6.14.0 et les notes de mise à jour

Date	Événement	Description
02/11/2023	Déploiement terminé	Amazon EMR 6.14.0 entièrement déployé dans toutes les régions prises en charge
10/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 6.14.0 publiées pour la première fois
04/10/2023	Première version	Amazon EMR 6.14.0 a été déployé pour la première fois dans les régions commerciales initiales

Amazon EMR version 6.13.0

Versions d'application 6.13.0

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
Kit AWS SDK pour Java	1,12,513	1,12,490	1,12,446	1,12,446
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.4.0	2.2.0	2.2.0
Flink	1.17.0	1.17.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4,17	2.4,17	2.4.15	2.4.15

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-1	0.13.1-amzn-0	0.13.0-amzn-0	0.13.0-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.11.0
Iceberg	1.3.0-amzn-1	1.3.0-amzn-0	1.2.0-amzn-0	1.2.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.3	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,281	0,281	0,279	0,279
Spark	3.4.1	3.4.0	3.3.2	3.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,1,2	0,1,2	0,1,2	0,1,2

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
Trino (PrestoSQL)	414	414	410	410
Zeppelin	0,1,1	0,1,1	0,1,1	0,1,1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notes de mise à jour 6.13.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.13.0 d'Amazon. Les modifications ont été apportées à la version 6.12.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.13.0](#).

Nouvelles fonctionnalités

- Amazon EMR 6.13.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-1, CUDA Toolkit 11.8.0, and JupyterHub 1.5.0.

Modifications, améliorations et problèmes résolus

- La version 6.13.0 améliore le démon de gestion des EMR journaux Amazon afin de garantir que tous les journaux sont chargés à une cadence régulière vers Amazon S3 lorsqu'une commande de fin de cluster est émise. Cela permet de résilier plus rapidement les clusters.
- La version 6.13.0 améliore les fonctionnalités de gestion des EMR journaux d'Amazon afin de garantir un téléchargement cohérent et rapide de tous les fichiers journaux vers Amazon S3. Cela profite particulièrement aux EMR clusters de longue durée.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)
2,0.2023 4,1,4322 906.0		4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

Versions Java par défaut 6.13.0

EMR Les versions 6.12.0 et ultérieures d'Amazon prennent en charge toutes les applications dotées d'Amazon Corretto 8 par défaut, à l'exception de Trino. Pour Trino, Amazon EMR prend en charge Amazon Corretto 17 par défaut à partir de la version 6.9.0 d'Amazon. EMR Amazon prend également en charge certaines applications avec Amazon Corretto 11 et 17. Ces applications sont répertoriées dans le tableau suivant. Si vous souhaitez modifier la valeur par défaut JVM de votre cluster, suivez les instructions fournies [Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique](#) pour chaque application exécutée sur le cluster. Vous ne pouvez utiliser qu'une seule version d'exécution de Java pour un cluster. Amazon EMR ne prend pas en charge l'exécution de différents nœuds ou applications sur différentes versions d'exécution sur le même cluster.

Bien qu'Amazon EMR prenne en charge Amazon Corretto 11 et 17 sur Apache Spark, Apache Hadoop et Apache Hive, les performances peuvent régresser pour certaines charges de travail lorsque vous utilisez ces versions de Corretto. Nous vous recommandons de tester vos charges de travail avant de modifier les valeurs par défaut.

Le tableau suivant indique les versions Java par défaut pour les applications dans Amazon EMR 6.13.0 :

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Étincelle RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versions des composants 6.13.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion`

commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.4.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
delta-standalone-connectors	0.6.0	Les connecteurs Delta fournissent différents environnements d'exécution pour intégrer Delta Lake avec des moteurs tels que Flink, Hive et Presto.
emr-ddb	5.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.6.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.10.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise

Composant	Version	Description
emr-s3-dist-cp	2.27.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.6.0	EMRConnecteur S3Select
emr-wal-cli	1.1.0	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,58,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.17.0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.17.0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	3.3.3-amzn-5	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.3-amzn-5	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.3-amzn-5	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.3-amzn-5	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.3-amzn-5	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httfs-server	3.3.3-amzn-5	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.3-amzn-5	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-5	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-5	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	3.3.3-amzn-5	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.3-amzn-5	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.17-amzn-1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.17-amzn-1	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.17-amzn-1	HBaseclient en ligne de commande.
hbase-rest-server	2.4.17-amzn-1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.17-amzn-1	Service fournissant un point de terminaison Thrift pour HBase
hbase-operator-tools	2.4.17-amzn-1	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-6	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	3.1.3-amzn-6	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-6	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-6	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-6	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-6	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-6	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.13.1-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.13.1-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.13.1-amzn-1	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.

Composant	Version	Description
hudi-spark	0.13.1-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.3.0-amzn-1	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.5.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.7.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.3	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.3	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.3	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.281-amzn-1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.281-amzn-1	Service pour exécuter les éléments d'une requête.
presto-client	0.281-amzn-1	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	414-amzn-1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.

Composant	Version	Description
trino-worker	414-amzn-1	Service pour exécuter les éléments d'une requête.
trino-client	414-amzn-1	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.4.1-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.4.1-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.4.1-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.4.1-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	23.06.0-amzn-1	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.

Composant	Version	Description
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-4	L'YARN application Tez et les bibliothèques.
tez-on-worker	0.10.2-amzn-4	L'YARN application tez et les bibliothèques pour les nœuds de travail.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.13.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.13.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Modifier la page d'accueil Java de KMS Hadoop	Not available.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.

Classifications	Description	Actions de reconfiguration
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.

Classifications	Description	Actions de reconfiguration
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier <code>lakeformation.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-delta	Modification des valeurs dans le fichier <code>delta.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier <code>exchange-manager.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier <code>iceberg.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifiez les valeurs dans le fichier <code>hudi.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.13.0

Journal des modifications pour la version 6.13.0 et les notes de mise à jour

Date	Événement	Description
23/09/2023	Déploiement terminé	Amazon EMR 6.13.0 entièrement déployé dans toutes les régions prises en charge
12/09/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 6.13.0 publiées pour la première fois
01/09/2023	Première version	Amazon EMR 6.13.0 a été déployé pour la première fois dans les régions commerciales initiales

Amazon EMR version 6.12.0

Versions d'application 6.12.0

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
Kit AWS SDK pour Java	1,12,490	1,12,446	1,12,446	1,12,397
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.4.0	2.2.0	2.2.0	2.2.0
Flink	1.17.0	1.16.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4,17	2.4.15	2.4.15	2.4.15

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.1-amzn-0	0.13.0-amzn-0	0.13.0-amzn-0	0.12.2-amzn-0
Hue	4.11.0	4.11.0	4.11.0	4.10.0
Iceberg	1.3.0-amzn-0	1.2.0-amzn-0	1.2.0-amzn-0	1.1.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.3	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,281	0,279	0,279	0,278
Spark	3.4.0	3.3.2	3.3.2	3.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,1,2	0,1,2	0,1,2	0,1,2

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
Trino (PrestoSQL)	414	410	410	403
Zeppelin	0,1,1	0,1,1	0,1,1	0,1,1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notes de mise à jour 6.12.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.12.0 d'Amazon. Les modifications ont été apportées à la version 6.11.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.12.0](#).

Nouvelles fonctionnalités

- Amazon EMR 6.12.0 supports Apache Spark 3.4.0, Apache Spark RAPIDS 23.06.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.1-amzn-0, Apache Iceberg 1.3.0-amzn-0, Trino 414, and PrestoDB 0.281.
- EMR Les versions 6.12.0 et supérieures d'Amazon prennent en charge LDAP l'intégration avec Apache Livy, Apache Hive via HiveServer 2 (HS2), Trino, Presto et Hue. Vous pouvez également installer Apache Spark et Apache Hadoop sur un EMR cluster utilisant la version 6.12.0 ou une version ultérieure et les configurer pour les utiliser. LDAP Pour plus d'informations, consultez [Utiliser Active Directory ou LDAP des serveurs pour l'authentification auprès d'Amazon EMR](#).

Modifications, améliorations et problèmes résolus

- EMR Les versions 6.12.0 et supérieures d'Amazon fournissent un support d'exécution Java 11 pour Flink. Pour plus d'informations, consultez [Configurer Flink pour qu'il fonctionne avec Java 11](#).
- La version 6.12.0 ajoute un nouveau mécanisme de nouvelle tentative au flux de travail de dimensionnement des EMR clusters pour les clusters qui exécutent Presto ou Trino. Cette amélioration réduit le risque que le redimensionnement du cluster soit bloqué indéfiniment en raison de l'échec d'une seule opération de redimensionnement. Cela améliore également l'utilisation du cluster, car celui-ci augmente et diminue la capacité plus rapidement.

- La version 6.12.0 corrige un problème à cause duquel les opérations de réduction de la taille du cluster peuvent se bloquer lorsqu'un nœud principal en cours de mise hors service appropriée tombe en panne pour une raison quelconque avant d'être complètement mis hors service.
- La version 6.12.0 améliore la logique de réduction du cluster afin que celui-ci ne tente pas de réduire les nœuds principaux en dessous du paramètre de facteur de HDFS réplication défini pour le cluster. Cela répond à vos exigences en matière de redondance des données et réduit le risque de blocage d'une opération de dimensionnement.
- La version 6.12.0 améliore les performances et l'efficacité du service de surveillance de l'état pour Amazon EMR en augmentant la vitesse à laquelle il enregistre les changements d'état des instances. Cette amélioration réduit le risque de dégradation des performances des nœuds de cluster qui exécutent plusieurs outils clients personnalisés ou des applications tierces.
- La version 6.12.0 améliore les performances du démon de gestion des journaux sur le cluster pour Amazon EMR. Par conséquent, les risques de dégradation des performances sont moindres avec les EMR clusters qui exécutent des étapes avec une grande simultanéité.
- Avec la EMR version 6.12.0 d'Amazon, le démon de gestion des journaux a été mis à niveau pour identifier tous les journaux utilisés activement avec des descripteurs de fichiers ouverts sur le stockage d'instance local, ainsi que les processus associés. Cette mise à niveau garantit qu'Amazon supprime EMR correctement les fichiers et récupère de l'espace de stockage une fois les journaux archivés dans Amazon S3.
- La version 6.12.0 inclut une amélioration du démon de gestion des journaux qui supprime les répertoires d'étapes vides et inutilisés dans le système de fichiers du cluster local. Un trop grand nombre de répertoires vides peut dégrader les performances des EMR démons Amazon et entraîner une surutilisation des disques.
- La version 6.12.0 permet la rotation des journaux pour les journaux de YARN Timeline Server. Cela permet de minimiser les scénarios de surutilisation des disques, en particulier pour les clusters de longue durée.
- La taille du volume racine par défaut est passée à 15 Go dans Amazon EMR 6.10.0 et versions ultérieures. Les versions antérieures ont une taille de volume racine par défaut de 10 Go.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023.0926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)
2,0.2023 4,1,4322 906.0		4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre)

Versions Java par défaut 6.12.0

EMR Les versions 6.12.0 et ultérieures d'Amazon prennent en charge toutes les applications dotées d'Amazon Corretto 8 par défaut, à l'exception de Trino. Pour Trino, Amazon EMR prend en charge Amazon Corretto 17 par défaut à partir de la version 6.9.0 d'Amazon. EMR Amazon prend également en charge certaines applications avec Amazon Corretto 11 et 17. Ces applications sont répertoriées dans le tableau suivant. Si vous souhaitez modifier la valeur par défaut JVM de votre cluster, suivez les instructions fournies [Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique](#) pour chaque application exécutée sur le cluster. Vous ne pouvez utiliser qu'une seule version d'exécution de Java pour un cluster. Amazon EMR ne prend pas en charge l'exécution de différents nœuds ou applications sur différentes versions d'exécution sur le même cluster.

Bien qu'Amazon EMR prenne en charge Amazon Corretto 11 et 17 sur Apache Spark, Apache Hadoop et Apache Hive, les performances peuvent régresser pour certaines charges de travail lorsque vous utilisez ces versions de Corretto. Nous vous recommandons de tester vos charges de travail avant de modifier les valeurs par défaut.

Le tableau suivant indique les versions Java par défaut pour les applications dans Amazon EMR 6.12.0 :

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8
Hudi	17, 11, 8

Application	Version Java/Amazon Corretto (la valeur par défaut est en gras)
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Étincelle RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versions des composants 6.12.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion`

commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.4.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
delta-standalone-connectors	0.6.0	Les connecteurs Delta fournissent différents environnements d'exécution pour intégrer Delta Lake avec des moteurs tels que Flink, Hive et Presto.
emr-ddb	5.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.5.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.9.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise

Composant	Version	Description
emr-s3-dist-cp	2.26.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.5.0	EMRConnecteur S3Select
emr-wal-cli	1.1.0	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,57,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.17.0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.17.0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	3.3.3-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.3-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.3-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.3-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.3-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httfs-server	3.3.3-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.3-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	3.3.3-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.3-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.17-amzn-0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.17-amzn-0	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.17-amzn-0	HBaseclient en ligne de commande.
hbase-rest-server	2.4.17-amzn-0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.17-amzn-0	Service fournissant un point de terminaison Thrift pour HBase
hbase-operator-tools	2.4.17-amzn-0	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-5	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	3.1.3-amzn-5	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-5	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-5	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-5	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-5	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-5	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.13.1-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.13.1-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.13.1-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.

Composant	Version	Description
hudi-spark	0.13.1-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.3.0-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.7.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.3	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.3	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.3	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.281-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.281-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.281-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	414-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.

Composant	Version	Description
trino-worker	414-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	414-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.4.0-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.4.0-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.4.0-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.4.0-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	23.06.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.

Composant	Version	Description
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-3	L'YARN application Tez et les bibliothèques.
tez-on-worker	0.10.2-amzn-3	L'YARN application tez et les bibliothèques pour les nœuds de travail.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.12.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.12.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Modifier la page d'accueil Java de KMS Hadoop	Not available.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.

Classifications	Description	Actions de reconfiguration
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.

Classifications	Description	Actions de reconfiguration
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier <code>lakeformation.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-delta	Modification des valeurs dans le fichier <code>delta.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier <code>exchange-manager.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier <code>iceberg.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifiez les valeurs dans le fichier <code>hudi.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.12.0

Journal des modifications pour la version 6.12.0 et les notes de mise à jour

Date	Événement	Description
27/07/2023	Documentation mise à jour	Mettez à jour les options Java pour la version 6.12 et ajoutez le didacticiel Oozie pour mettre à jour JVM
21/07/2023	Déploiement terminé	Amazon EMR 6.12.0 entièrement déployé dans toutes les régions prises en charge
21/07/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 6.12.0 publiées pour la première fois
12/07/2023	Première version	Amazon EMR 6.12.0 a été déployé pour la première fois dans les régions commerciales initiales

Amazon EMR version 6.11.1

Versions d'application 6.11.1

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
Kit AWS SDK pour Java	1,12,446	1,12,446	1,12,397	1,12,397
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.2.0	2.2.0
Flink	1.16.0	1.16.0	1.16.0	1.16.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.15	2.4.15

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.0-amzn-0	0.13.0-amzn-0	0.12.2-amzn-0	0.12.2-amzn-0
Hue	4.11.0	4.11.0	4.10.0	4.10.0
Iceberg	1.2.0-amzn-0	1.2.0-amzn-0	1.1.0-amzn-0	1.1.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.4.1	1.5.0	1.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,279	0,279	0,278	0,278
Spark	3.3.2	3.3.2	3.3.1	3.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.11.0
Tez	0,1,2	0,1,2	0,1,2	0,1,2

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
Trino (PrestoSQL)	410	410	403	403
Zeppelin	0,1,1	0,1,1	0,1,1	0,1,1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notes de mise à jour 6.11.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.11.1 d'Amazon. Les modifications ont été apportées à la version 6.11.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.11.1](#).

Modifications, améliorations et problèmes résolus

- En raison d'un conflit de verrouillage, un nœud peut se retrouver bloqué s'il est ajouté ou supprimé en même temps qu'il est mis hors service. Par conséquent, le gestionnaire de ressources Hadoop (YARN) ne répond plus et affecte tous les conteneurs entrants et en cours d'exécution.
- Cette version inclut une modification qui permet aux clusters à haute disponibilité de se remettre d'un état défaillant après le redémarrage.
- Cette version inclut des correctifs de sécurité pour Hue etHBase.
- Cette version résout un problème selon lequel les clusters exécutant des charges de travail sur Spark avec Amazon EMR peuvent recevoir silencieusement des résultats incorrects avec `contains`, `startsWith`, `endsWith`, et `like`. Ce problème se produit lorsque vous utilisez les expressions sur des champs partitionnés contenant des métadonnées sur le serveur Amazon EMR Hive3 Metastore (`hive.metastore`). HMS
- Cette version corrige un problème de régulation du côté de Glue lorsqu'il n'existe aucune fonction définie par l'utilisateur (`glue.udf`). UDF
- Cette version corrige un problème qui supprime les journaux des conteneurs par le service d'agrégation des journaux des nœuds avant que le lanceur de journaux ne puisse les envoyer vers S3 en cas de YARN mise hors service.
- Cette version corrige un problème lié aux métriques du FairShare planificateur lorsque Node Label est activé pour Hadoop.

- Cette version corrige un problème qui affectait les performances de Spark lorsque vous définissez une valeur `true` par défaut pour la configuration `spark.yarn.heterogeneousExecutors.enabled` dans `spark-defaults.conf`.
- Cette version corrige un problème lié à l'échec de la lecture des données de shuffle par Reduce Task. Ce problème provoquait des échecs de requêtes Hive avec une erreur de mémoire corrompue.
- Cette version ajoute un nouveau mécanisme de nouvelle tentative au flux de travail de dimensionnement des EMR clusters pour les clusters qui exécutent Presto ou Trino. Cette amélioration réduit le risque que le redimensionnement du cluster soit bloqué indéfiniment en raison de l'échec d'une seule opération de redimensionnement. Cela améliore également l'utilisation du cluster, car celui-ci augmente et diminue la capacité plus rapidement.
- Cette version améliore la logique de réduction du cluster afin que votre cluster ne tente pas de réduire les nœuds principaux en dessous du paramètre de facteur de HDFS réplication défini pour le cluster. Cela répond à vos exigences en matière de redondance des données et réduit le risque de blocage d'une opération de dimensionnement.
- Le démon de gestion des journaux a été mis à niveau pour identifier tous les journaux en cours d'utilisation avec des descripteurs de fichiers ouverts sur le stockage d'instance local, ainsi que les processus associés. Cette mise à niveau garantit qu'Amazon supprime EMR correctement les fichiers et récupère de l'espace de stockage une fois les journaux archivés dans Amazon S3.
- Cette version inclut une amélioration du démon de gestion des journaux qui supprime les répertoires d'étapes vides et inutilisés dans le système de fichiers du cluster local. Un trop grand nombre de répertoires vides peut dégrader les performances des EMR démons Amazon et entraîner une surutilisation des disques.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelea Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0_20230727.0	4.14.320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

Versions des composants 6.11.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.2.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
delta-standalone-connectors	0.6.0	Les connecteurs Delta fournissent différents environnements d'exécution pour intégrer Delta Lake avec des moteurs tels que Flink, Hive et Presto.
emr-ddb	5.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.8.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.25.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.4.0	EMRConnecteur S3Select
emr-wal-cli	1.1.0	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,56,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.16.0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.16.0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.3.3-amzn-3.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.3-amzn-3.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.3-amzn-3.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.3-amzn-3.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.3-amzn-3.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.3.3-amzn-3.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.3-amzn-3.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-3.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	3.3.3-amzn-3.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.3-amzn-3.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.3-amzn-3.1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.15-amzn-1.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.15-amzn-1.1	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.15-amzn-1.1	HBaseclient en ligne de commande.
hbase-rest-server	2.4.15-amzn-1.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.15-amzn-1.1	Service fournissant un point de terminaison Thrift pour. HBase
hbase-operator-tools	2.4.15-amzn-1.1	Outil de réparation pour les HBase clusters Apache.

Composant	Version	Description
hcatalog-client	3.1.3-amzn-4.1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-4.1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-4.1	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-4.1	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-4.1	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-4.1	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-4.1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.13.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.13.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.

Composant	Version	Description
hudi-trino	0.13.0-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.13.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.2.0-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda

Composant	Version	Description
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.2	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.2	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.279-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.279-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.279-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
trino-coordinator	410-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	410-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	410-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.3.2-amzn-0.1	Clients de ligne de commande Spark.
spark-history-server	3.3.2-amzn-0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.3.2-amzn-0.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.3.2-amzn-0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
spark-rapids	23.02.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-2.1	L'YARNapplication Tez et les bibliothèques.
tez-on-worker	0.10.2-amzn-2.1	L'YARNapplication tez et les bibliothèques pour les nœuds de travail.
webserver	2.4.41+	HTTPserveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.11.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.11.1

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the Resource Manager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop

Classifications	Description	Actions de reconfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.

Classifications	Description	Actions de reconfiguration
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.

Classifications	Description	Actions de reconfiguration
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier lakeformation.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-delta	Modification des valeurs dans le fichier delta.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier exchange-manager.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier iceberg.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifiez les valeurs dans le fichier hudi.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.

Classifications	Description	Actions de reconfiguration
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.11.1

Journal des modifications pour la version 6.11.1 et les notes de mise à jour

Date	Événement	Description
30/08/2023	Mise à jour des notes de mise à jour	Ajout de plusieurs corrections liées au plan de contrôle dans les notes de mise à jour
21/08/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 6.11.1 publiées pour la première fois
16/08/2023	Déploiement terminé	Amazon EMR 6.11.1 entièrement déployé dans toutes les régions prises en charge
04/08/2023	Première version	Amazon EMR 6.11.1 a été déployé pour la première fois dans des régions commerciales limitées

Amazon EMR version 6.11.0

Versions d'application 6.11.0

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
Kit AWS SDK pour Java	1,12,446	1,12,397	1,12,397	1,12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.2.0	2.1.0
Flink	1.16.0	1.16.0	1.16.0	1.15.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.15	2.4.13

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.13.0-amzn-0	0.12.2-amzn-0	0.12.2-amzn-0	0.12.1-amzn-0
Hue	4.11.0	4.10.0	4.10.0	4.10.0
Iceberg	1.2.0-amzn-0	1.1.0-amzn-0	1.1.0-amzn-0	0.14.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.4.1	1.5.0	1.5.0	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,279	0,278	0,278	0,276
Spark	3.3.2	3.3.1	3.3.1	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.11.0	2.10.0
Tez	0,1,2	0,1,2	0,1,2	0,1,2

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
Trino (PrestoSQL)	410	403	403	398
Zeppelin	0,1,1	0,1,1	0,1,1	0,1,1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notes de mise à jour 6.11.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.11.0 d'Amazon. Les modifications ont été apportées à la version 6.10.0. Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Nouvelles fonctionnalités

- Amazon EMR 6.11.0 prend en charge Apache Spark 3.3.2-amzn-0, Apache Spark RAPIDS 23.02.0-amzn-0, 11.8.0, Apache Hudi 0.13.0-amzn-0, Apache Iceberg CUDA 1.2.0-amzn-0, Trino 410-amzn-0 et PrestoDB 0.279-amzn-0.

Modifications, améliorations et problèmes résolus

- Avec Amazon EMR 6.11.0, le connecteur DynamoDB a été mis à niveau vers la version 5.0.0. La version 5.0.0 utilise AWS SDK for Java 2.x. Les versions précédentes utilisaient la version AWS SDK for Java 1.x. À la suite de cette mise à niveau, nous vous conseillons vivement de tester votre code avant d'utiliser le connecteur DynamoDB avec Amazon 6.11. EMR
- Lorsque le connecteur DynamoDB pour EMR Amazon 6.11.0 appelle le service DynamoDB, il utilise la valeur de région que vous fournissez pour la propriété `dynamodb.endpoint`. Nous vous recommandons de configurer également `dynamodb.region` lorsque vous utilisez `dynamodb.endpoint`, et que les deux propriétés ciblent la même Région AWS. Si vous l'utilisez `dynamodb.endpoint` et que vous ne le configurez pas `dynamodb.region`, le connecteur DynamoDB pour EMR Amazon 6.11.0 renverra une exception de région non valide et tentera de rapprocher vos Région AWS informations provenant du service de métadonnées d'instance EC2 Amazon (). IMDS Si le connecteur ne parvient pas à récupérer la région IMDS, la valeur par défaut est des États-Unis (Virginie du Nord) (`us-east-1`). L'erreur suivante illustre l'exception de région non valide que vous pourriez

obtenir si vous ne configurez pas correctement la propriété `dynamodb.region` : `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region`. Pour plus d'informations sur les classes concernées par la AWS SDK for Java mise à niveau vers la version 2.x, consultez le commit [Upgrade AWS SDK for Java from 1.x to 2.x \(#175\)](#) dans le GitHub dépôt du connecteur Amazon - EMR DynamoDB.

- Cette version corrige un problème où les données des colonnes deviennent NULL lorsque vous utilisez Delta Lake pour stocker les données de la table Delta dans Amazon S3 après l'opération de changement de nom de colonne. Pour plus d'informations sur cette fonctionnalité expérimentale dans Delta Lake, consultez [Opération de changement de nom de colonne](#) dans le guide de l'utilisateur Delta Lake.
- La version 6.11.0 résout un problème qui peut survenir lorsque vous créez un nœud périphérique en répliquant l'un des nœuds primaires à partir d'un cluster comportant plusieurs nœuds primaires. Le nœud périphérique répliqué peut retarder les opérations de réduction d'échelle ou entraîner une utilisation élevée de la mémoire sur les nœuds primaires. Pour plus d'informations sur la création d'un nœud périphérique pour communiquer avec votre EMR cluster, consultez la section [Edge Node Creator](#) dans le `aws-samples` référentiel sur GitHub.
- La version 6.11.0 améliore le processus d'automatisation EMR utilisé par Amazon pour remonter les volumes EBS Amazon sur une instance après un redémarrage.
- La version 6.11.0 corrige un problème qui entraînait des écarts intermittents dans les métriques Hadoop publiées par Amazon sur Amazon. EMR CloudWatch
- La version 6.11.0 résout un problème lié aux EMR clusters où une mise à jour du fichier de YARN configuration contenant la liste d'exclusion des nœuds du cluster est interrompue en raison d'une surutilisation du disque. La mise à jour incomplète entrave les futures opérations de réduction de la taille du cluster. Cette version garantit que votre cluster reste sain et que les opérations de dimensionnement fonctionnent comme prévu.
- La taille du volume racine par défaut est passée à 15 Go dans Amazon EMR 6.10.0 et versions ultérieures. Les versions antérieures ont une taille de volume racine par défaut de 10 Go.
- Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans activation de la mise à l'échelle gérée.

Dans les EMR versions 6.11.0 et supérieures d'Amazon ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` est définie sur `false` in `yarn-site.xml` pour résoudre ce problème.

Bien que le correctif corrige les problèmes introduits par YARN -9608, il peut entraîner l'échec des tâches Hive en raison de la perte de données par shuffle sur les clusters sur lesquels le dimensionnement géré est activé. Nous avons atténué ce risque dans cette version en définissant également `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` pour les charges de travail Hive. Cette configuration n'est disponible qu'avec les EMR versions 6.11.0 et supérieures d'Amazon.

- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

Note

Cette version ne reçoit plus de AMI mises à jour automatiques puisqu'elle a été remplacée par une autre version de correctif. La version du correctif est indiquée par le numéro qui suit la deuxième décimale (6.8.1). Pour savoir si vous utilisez la dernière version du correctif, consultez les versions disponibles dans le [guide des versions](#), ou consultez le menu déroulant des EMRversions d'Amazon lorsque vous créez un cluster dans la console, ou utilisez l'[list-release-labels](#) CLI action [ListReleaseLabels](#) API ou. Pour être tenu au courant des nouvelles versions, abonnez-vous au RSS fil d'actualité sur le [site Quoi de neuf ?](#) page.

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023.0808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre)

Versions des composants 6.11.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.2.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
<code>delta-standalone-connectors</code>	0.6.0	Les connecteurs Delta fournissent différents environnements d'exécution pour intégrer Delta Lake avec des moteurs tels que Flink, Hive et Presto.
<code>emr-ddb</code>	5.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-goodies	3.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.8.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.25.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.4.0	EMRConnecteur S3Select
emr-wal-cli	1.1.0	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,56,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.16.0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.16.0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.3.3-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.3-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.3-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.3-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.3-amzn-3	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	3.3.3-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.3-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	3.3.3-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.3-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.3-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.15-amzn-1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.15-amzn-1	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.15-amzn-1	HBaseclient en ligne de commande.
hbase-rest-server	2.4.15-amzn-1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.15-amzn-1	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hbase-operator-tools	2.4.15-amzn-1	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-4	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-4	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-4	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-4	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-4	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-4	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-4	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.13.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.

Composant	Version	Description
hudi-presto	0.13.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.13.0-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.13.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.2.0-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.

Composant	Version	Description
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.2	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.2	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.279-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.279-amzn-0	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
presto-client	0.279-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	410-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	410-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	410-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.3.2-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.3.2-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	3.3.2-amzn-0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.3.2-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	23.02.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-2	L'YARNapplication Tez et les bibliothèques.
tez-on-worker	0.10.2-amzn-2	L'YARNapplication tez et les bibliothèques pour les nœuds de travail.
webserver	2.4.41+	HTTPserveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.11.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.11.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the Resource Manager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, HiveMetaStore, and MapReduceHistoryServer.

Classifications	Description	Actions de reconfiguration
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Actions de reconfiguration
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.

Classifications	Description	Actions de reconfiguration
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.

Classifications	Description	Actions de reconfiguration
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-lakeformat ion	Modifiez les valeurs dans le fichier <code>lakeformation.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)

Classifications	Description	Actions de reconfiguration
trino-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-delta	Modification des valeurs dans le fichier <code>delta.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier <code>exchange-manager.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier <code>iceberg.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifiez les valeurs dans le fichier <code>hudi.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.11.0

Journal des modifications pour la version 6.11.0 et les notes de mise à jour

Date	Événement	Description
21/08/2023	Mettre à jour	Correction d'un problème introduit avec Hadoop 3.3.3.
26/07/2023	Mettre à jour	Étiquettes 2.0.20230612.0 et 2.0.20230628.0 de la nouvelle version du système d'exploitation.

Date	Événement	Description
09/06/2023	Déploiement terminé	Amazon EMR 6.11.0 entièrement déployé dans toutes les régions prises en charge
09/06/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 6.11.0 publiées pour la première fois
08/06/2023	Première version	Amazon EMR 6.11.0 a été déployé pour la première fois dans les régions commerciales initiales

Amazon EMR version 6.10.1

Versions d'application 6.10.1

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
Kit AWS SDK pour Java	1,12,397	1,12,397	1,12,170	1,12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.2.0	2.1.0	2.1.0
Flink	1.16.0	1.16.0	1.15.2	1.15.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.15	2.4.13	2.4.13
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.3.3
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.2-amzn-0	0.12.2-amzn-0	0.12.1-amzn-0	0.12.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	1.1.0-amzn-0	1.1.0-amzn-0	0.14.1-amzn-0	0.14.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.6.0
JupyterHub	1.5.0	1.5.0	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,278	0,278	0,276	0,276
Spark	3.3.1	3.3.1	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.11.0	2.10.0	2.10.0
Tez	0,1,2	0,1,2	0,1,2	0,1,2
Trino (PrestoSQL)	403	403	398	398
Zeppelin	0,1,1	0,1,1	0,1,1	0,1,1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notes de mise à jour 6.10.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.10.1 d'Amazon. Les modifications ont été apportées à la version 6.10.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.10.1](#).

Modifications, améliorations et problèmes résolus

- En raison d'un conflit de verrouillage, un nœud peut se retrouver bloqué s'il est ajouté ou supprimé en même temps qu'il est mis hors service. Par conséquent, le gestionnaire de ressources Hadoop (YARN) ne répond plus et affecte tous les conteneurs entrants et en cours d'exécution.
- Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application

soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans activation de la mise à l'échelle gérée.

Dans les EMR versions 6.11.0 et supérieures d'Amazon ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` est définie sur `false` in `yarn-site.xml` pour résoudre ce problème.

Bien que le correctif corrige les problèmes introduits par YARN -9608, il peut entraîner l'échec des tâches Hive en raison de la perte de données par shuffle sur les clusters sur lesquels le dimensionnement géré est activé. Nous avons atténué ce risque dans cette version en définissant également `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` pour les charges de travail Hive. Cette configuration n'est disponible qu'avec les EMR versions 6.11.0 et supérieures d'Amazon.

- Le collecteur de métriques n'enverra aucune métrique au plan de contrôle après le basculement du nœud primaire dans les clusters avec la configuration des groupes d'instances.
- Cette version inclut une modification qui permet aux clusters à haute disponibilité de se remettre d'un état défaillant après le redémarrage.
- Cette version inclut des correctifs de sécurité pour Hue et HBase.
- Cette version résout un problème selon lequel les clusters exécutant des charges de travail sur Spark avec Amazon EMR peuvent recevoir silencieusement des résultats incorrects avec `contains`, `startsWith`, `endsWith`, et `like`. Ce problème se produit lorsque vous utilisez les expressions sur des champs partitionnés contenant des métadonnées sur le serveur Amazon EMR Hive3 Metastore (). HMS
- Cette version corrige un problème de régulation du côté de Glue lorsqu'il n'existe aucune fonction définie par l'utilisateur (). UDF
- Cette version corrige un problème qui supprime les journaux des conteneurs par le service d'agrégation des journaux des nœuds avant que le lanceur de journaux ne puisse les envoyer vers S3 en cas de YARN mise hors service.
- Cette version corrige un problème lié aux métriques du FairShare planificateur lorsque Node Label est activé pour Hadoop.
- Cette version corrige un problème qui affectait les performances de Spark lorsque vous définissez une valeur `true` par défaut pour la configuration `spark.yarn.heterogeneousExecutors.enabled` dans `spark-defaults.conf`.

- Cette version corrige un problème lié à l'échec de la lecture des données de shuffle par Reduce Task. Ce problème provoquait des échecs de requêtes Hive avec une erreur de mémoire corrompue.
- Cette version ajoute un nouveau mécanisme de nouvelle tentative au flux de travail de dimensionnement des EMR clusters pour les clusters qui exécutent Presto ou Trino. Cette amélioration réduit le risque que le redimensionnement du cluster soit bloqué indéfiniment en raison de l'échec d'une seule opération de redimensionnement. Cela améliore également l'utilisation du cluster, car celui-ci augmente et diminue la capacité plus rapidement.
- Cette version améliore la logique de réduction du cluster afin que votre cluster ne tente pas de réduire les nœuds principaux en dessous du paramètre de facteur de HDFS réplication défini pour le cluster. Cela répond à vos exigences en matière de redondance des données et réduit le risque de blocage d'une opération de dimensionnement.
- Le démon de gestion des journaux a été mis à niveau pour identifier tous les journaux en cours d'utilisation avec des descripteurs de fichiers ouverts sur le stockage d'instance local, ainsi que les processus associés. Cette mise à niveau garantit qu'Amazon supprime EMR correctement les fichiers et récupère de l'espace de stockage une fois les journaux archivés dans Amazon S3.
- Cette version inclut une amélioration du démon de gestion des journaux qui supprime les répertoires d'étapes vides et inutilisés dans le système de fichiers du cluster local. Un trop grand nombre de répertoires vides peut dégrader les performances des EMR démons Amazon et entraîner une surutilisation des disques.
- Cette version résout un problème qui peut survenir lorsque vous créez un nœud périphérique en répliquant l'un des nœuds primaires à partir d'un cluster comportant plusieurs nœuds primaires. Le nœud périphérique répliqué peut retarder les opérations de réduction d'échelle ou entraîner une utilisation élevée de la mémoire sur les nœuds primaires. Pour plus d'informations sur la création d'un nœud périphérique pour communiquer avec votre EMR cluster, consultez la section [Edge Node Creator](#) dans le `aws-samples` référentiel sur GitHub.
- Cette version améliore le processus d'automatisation EMR utilisé par Amazon pour remonter les EBS volumes Amazon sur une instance après un redémarrage.
- Cette version corrige un problème qui entraînait des écarts intermittents dans les métriques Hadoop publiées par Amazon sur EMR Amazon. CloudWatch
- Cette version résout un problème lié aux EMR clusters où une mise à jour du fichier de YARN configuration contenant la liste d'exclusion des nœuds du cluster est interrompue en raison d'une surutilisation du disque. La mise à jour incomplète entrave les futures opérations de réduction de la taille du cluster. Cette version garantit que votre cluster reste sain et que les opérations de dimensionnement fonctionnent comme prévu.

- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelea Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

Versions des composants 6.10.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.2.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.7.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.24.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.3.0	EMRConnecteur S3Select

Composant	Version	Description
emr-wal-cli	1.0.0	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,55,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.16.0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.16.0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.3.3-amzn-2.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.3-amzn-2.1	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	3.3.3-amzn-2.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.3-amzn-2.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.3-amzn-2.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.3.3-amzn-2.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.3-amzn-2.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-2.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-2.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.3-amzn-2.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.3-amzn-2.1	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	2.4.15-amzn-0.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.15-amzn-0.1	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.15-amzn-0.1	HBaseclient en ligne de commande.
hbase-rest-server	2.4.15-amzn-0.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.15-amzn-0.1	Service fournissant un point de terminaison Thrift pour HBase
hbase-operator-tools	2.4.15-amzn-0.1	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-3.1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-3.1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-3.1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-3.1	Client de ligne de commande Hive.

Composant	Version	Description
hive-hbase	3.1.3-amzn-3.1	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-3.1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-3.1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.12.2-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.12.2-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.12.2-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.12.2-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.1.0-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.

Composant	Version	Description
jupyterhub	1.5.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	6,0.0- SNAPSHOT	Connecteurs Apache Phoenix pour Spark-3

Composant	Version	Description
phoenix-query-server	6.0.0	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.278.1-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.278.1-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.278.1-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	403-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	403-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	403-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.

Composant	Version	Description
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.3.1-amzn-0.1	Clients de ligne de commande Spark.
spark-history-server	3.3.1-amzn-0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.3.1-amzn-0.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.3.1-amzn-0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	22.12.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-1.1	L'YARNapplication Tez et les bibliothèques.

Composant	Version	Description
tez-on-worker	0.10.2-amzn-1.1	L'YARN application tez et les bibliothèques pour les nœuds de travail.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.10.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.10.1

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classifications	Description	Actions de reconfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.

Classifications	Description	Actions de reconfiguration
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.

Classifications	Description	Actions de reconfiguration
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier lakeformation.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-delta	Modification des valeurs dans le fichier delta.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier exchange-manager.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier iceberg.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifiez les valeurs dans le fichier hudi.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.

Classifications	Description	Actions de reconfiguration
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.10.1

Journal des modifications pour la version 6.10.1 et les notes de mise à jour

Date	Événement	Description
30/08/2023	Mise à jour des notes de mise à jour	Ajout de plusieurs corrections liées au plan de contrôle dans les notes de mise à jour
21/08/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 6.10.1 publiées pour la première fois
16/08/2023	Déploiement terminé	Amazon EMR 6.10.1 entièrement déployé dans toutes les régions prises en charge
04/08/2023	Première version	Amazon EMR 6.10.1 a été déployé pour la première fois dans des régions commerciales limitées

Amazon EMR version 6.10.0

Versions d'application 6.10.0

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
Kit AWS SDK pour Java	1,12,397	1,12,170	1,12,170	1,12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.2.0	2.1.0	2.1.0	-
Flink	1.16.0	1.15.2	1.15.2	1.15.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.15	2.4.13	2.4.13	2.4.12

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.3.3	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.2-amzn-0	0.12.1-amzn-0	0.12.1-amzn-0	0.11.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	1.1.0-amzn-0	0.14.1-amzn-0	0.14.1-amzn-0	0.14.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.6.0	2.1.0
JupyterHub	1.5.0	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,278	0,276	0,276	0,273
Spark	3.3.1	3.3.0	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.11.0	2.10.0	2.10.0	2.9.1
Tez	0,1,2	0,1,2	0,1,2	0.9.2

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
Trino (PrestoSQL)	403	398	398	388
Zeppelin	0,1,1	0,1,1	0,1,1	0,1,1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notes de mise à jour 6.10.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.10.0 d'Amazon. Les modifications ont été apportées à la version 6.9.0. Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Nouvelles fonctionnalités

- Amazon EMR 6.10.0 prend en charge Apache Spark 3.3.1, Apache Spark RAPIDS 22.12.0, CUDA 11.8.0, Apache Hudi 0.12.2-amzn-0, Apache Iceberg 1.1.0-amzn-0, Trino 403 et PrestoDB 0.278.1.
- Amazon EMR 6.10.0 inclut un connecteur Trino-Hudi natif qui fournit un accès en lecture aux données des tables Hudi. Vous pouvez activer le connecteur avec `trino-cli --catalog hudi` et le configurer en fonction de vos besoins avec `trino-connector-hudi`. L'intégration native avec Amazon EMR signifie que vous n'avez plus besoin de l'utiliser `trino-connector-hive` pour interroger les tables Hudi. Pour obtenir la liste des configurations prises en charge avec le nouveau connecteur, consultez la page [Connecteur Hudi](#) dans la documentation de Trino.
- EMR Les versions 6.10.0 et supérieures d'Amazon prennent en charge l'intégration d'Apache Zeppelin avec Apache Flink. Pour plus d'informations, consultez [Travailler avec les jobs Flink de Zeppelin dans Amazon EMR](#).

Problèmes connus


- Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans activation de la mise à l'échelle gérée.

Pour contourner ce problème dans Amazon EMR 6.10.0, vous pouvez définir la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `false` in `yarn-site.xml`. Dans les EMR versions 6.11.0 et supérieures d'Amazon ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la configuration est définie `false` par défaut pour résoudre ce problème.

Modifications, améliorations et problèmes résolus

- Amazon EMR 6.10.0 supprime la dépendance à [l'intégration d'Amazon Redshift minimal-json.jar pour Apache Spark](#) et ajoute automatiquement les fichiers jar liés à Spark-Redshift requis au chemin de classe de l'exécuteur pour Spark :, et. `spark-redshift.jar` `spark-avro.jar` `RedshiftJDBC.jar`
- La version 6.10.0 améliore le démon de gestion des journaux sur le cluster afin de surveiller des dossiers de journaux supplémentaires dans votre cluster. EMR Cette amélioration permet de minimiser les scénarios de surutilisation des disques.
- La version 6.10.0 redémarre automatiquement le démon de gestion des journaux du cluster lorsqu'il s'arrête. Cette amélioration réduit le risque que les nœuds apparaissent défectueux en raison d'une surutilisation du disque.
- Amazon EMR 6.10.0 prend en charge les points de terminaison régionaux pour EMRFS le mappage des utilisateurs.
- La taille du volume racine par défaut est passée à 15 Go dans Amazon EMR 6.10.0 et versions ultérieures. Les versions antérieures ont une taille de volume racine par défaut de 10 Go.
- La version 6.10.0 corrige un problème qui provoquait le blocage des tâches Spark lorsque tous les exécuteurs Spark restants se trouvaient sur un hôte en cours de mise hors service avec le gestionnaire de ressources. YARN
- Avec Amazon EMR 6.6.0 à 6.9.x, les INSERT requêtes comportant une partition dynamique et une clause ORDER BY ou SORT BY auront toujours deux réducteurs. Ce problème est dû à la OSS modification [HIVE-20703](#), qui soumet l'optimisation des partitions de tri dynamique à une décision basée sur les coûts. Si votre charge de travail ne nécessite pas le tri des partitions dynamiques, nous vous recommandons de définir la propriété `hive.optimize.sort.dynamic.partition.threshold` sur `-1` pour désactiver la nouvelle fonctionnalité et obtenir le nombre de réducteurs correctement calculé. Ce problème est résolu dans OSS Hive dans le cadre de la version [HIVE-22269](#) et dans Amazon 6.10.0. EMR

- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

 Note

Cette version ne reçoit plus de AMI mises à jour automatiques puisqu'elle a été remplacée par une autre version de correctif. La version du correctif est indiquée par le numéro qui suit la deuxième décimale (6 . 8 . **1**). Pour savoir si vous utilisez la dernière version du correctif, consultez les versions disponibles dans le [guide des versions](#), ou consultez le menu déroulant des EMRversions d'Amazon lorsque vous créez un cluster dans la console, ou utilisez l'[list-release-labels](#) CLI action [ListReleaseLabels](#) API ou. Pour être tenu au courant des nouvelles versions, abonnez-vous au RSS fil d'actualité sur le [site Quoi de neuf ?](#) page.

OsRelea Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre), Israël (Tel Aviv)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong Kong), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong Kong), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 207.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong Kong), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE)

Versions des composants 6.10.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.2.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.7.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.24.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.3.0	EMRConnecteur S3Select

Composant	Version	Description
emr-wal-cli	1.0.0	CLI utilisée pour la liste/suppression EMR WAL.
emrfs	2,55,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.16.0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.16.0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.3.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	3.3.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.3.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.3-amzn-2	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.3.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	2.4.15-amzn-0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.15-amzn-0	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.15-amzn-0	HBaseclient en ligne de commande.
hbase-rest-server	2.4.15-amzn-0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.15-amzn-0	Service fournissant un point de terminaison Thrift pour HBase
hbase-operator-tools	2.4.15-amzn-0	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-3	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-3	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-3	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-3	Client de ligne de commande Hive.

Composant	Version	Description
hive-hbase	3.1.3-amzn-3	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-3	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-3	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.12.2-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.12.2-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.12.2-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.12.2-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	1.1.0-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.

Composant	Version	Description
jupyterhub	1.5.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,8.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	6,0.0- SNAPSHOT	Connecteurs Apache Phoenix pour Spark-3

Composant	Version	Description
phoenix-query-server	6.0.0	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.278.1-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.278.1-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.278.1-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	403-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	403-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	403-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.

Composant	Version	Description
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.3.1-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.3.1-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.3.1-amzn-0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.3.1-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	22.12.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-1	L'YARNapplication Tez et les bibliothèques.

Composant	Version	Description
tez-on-worker	0.10.2-amzn-1	L'YARN application tez et les bibliothèques pour les nœuds de travail.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.10.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.10.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classifications	Description	Actions de reconfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.

Classifications	Description	Actions de reconfiguration
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.

Classifications	Description	Actions de reconfiguration
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier lakeformation.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-delta	Modification des valeurs dans le fichier delta.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier exchange-manager.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier iceberg.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Modifiez les valeurs dans le fichier hudi.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.

Classifications	Description	Actions de reconfiguration
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Modifiez les valeurs dans le fichier kms-logback.xml de RangerKMS.	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.10.0

Journal des modifications pour la version 6.10.0 et les notes de mise à jour

Date	Événement	Description
21/08/2023	Mettre à jour	Ajout d'un problème connu introduit avec Hadoop 3.3.3.
26/07/2023	Mettre à jour	Étiquettes 2.0.20230612.0 et 2.0.20230628.0 de la nouvelle version du système d'exploitation.
03 avril	Déploiement terminé	Amazon EMR 6.10 entièrement déployé dans toutes les régions prises en charge
03 avril	Publication de documents	Notes de mise à jour d'Amazon EMR 6.10 publiées pour la première fois
27/02/2023	Première version	Amazon EMR 6.10 déployé dans un nombre limité de régions commerciales

Amazon EMR version 6.9.1

Versions d'application 6.9.1

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
Kit AWS SDK pour Java	1,12,170	1,12,170	1,12,170	1,12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.1.0	2.1.0	-	-
Flink	1.15.2	1.15.2	1.15.1	1.15.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.13	2.4.13	2.4.12	2.4.12

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.3.3	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.1-amzn-0	0.12.1-amzn-0	0.11.1-amzn-0	0.11.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.1-amzn-0	0.14.1-amzn-0	0.14.0-amzn-0	0.14.0-amzn-0
JupyterEnterpriseGateway	2.6.0	2.6.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.9.1
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,276	0,276	0,273	0,273
Spark	3.3.0	3.3.0	3.3.0	3.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.10.0	2.10.0	2.9.1	2.9.1
Tez	0,1,2	0,1,2	0.9.2	0.9.2

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
Trino (PrestoSQL)	398	398	388	388
Zeppelin	0,1,1	0,1,1	0,1,1	0,1,1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notes de mise à jour 6.9.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.9.1 d'Amazon. Les modifications ont été apportées à la version 6.9.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.9.1](#).

Modifications, améliorations et problèmes résolus

- Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans activation de la mise à l'échelle gérée.

Dans les EMR versions 6.11.0 et supérieures d'Amazon ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` est définie sur `false` in `yarn-site.xml` pour résoudre ce problème.

Bien que le correctif corrige les problèmes introduits par YARN -9608, il peut entraîner l'échec des tâches Hive en raison de la perte de données par shuffle sur les clusters sur lesquels le dimensionnement géré est activé. Nous avons atténué ce risque dans cette version en définissant également `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` pour les charges de travail Hive. Cette configuration n'est disponible qu'avec les EMR versions 6.11.0 et supérieures d'Amazon.

- Le collecteur de métriques n'enverra aucune métrique au plan de contrôle après le basculement du nœud primaire dans les clusters avec la configuration des groupes d'instances.
- Cette version élimine les tentatives en cas d'échec des HTTP demandes adressées aux points de terminaison du collecteur de mesures.

- Cette version inclut une modification qui permet aux clusters à haute disponibilité de se remettre d'un état défaillant après le redémarrage.
- Cette version corrige un problème selon lequel de grandes quantités créées par l'utilisateur UIDs provoquaient des exceptions de dépassement.
- Cette version corrige les problèmes de temporisation liés au processus de EMR reconfiguration d'Amazon.
- Cette version inclut des correctifs de sécurité.
- Cette version résout un problème selon lequel les clusters exécutant des charges de travail sur Spark avec Amazon EMR peuvent recevoir silencieusement des résultats incorrects avec `contains`, `startsWith`, `endsWith`, et `like`. Ce problème se produit lorsque vous utilisez les expressions sur des champs partitionnés contenant des métadonnées sur le serveur Amazon EMR Hive3 Metastore (`hive.metastore`). HMS
- Avec Amazon EMR 6.6.0 à 6.9.x, les INSERT requêtes comportant une partition dynamique et une clause ORDER BY ou SORT BY auront toujours deux réducteurs. Ce problème est dû à la OSS modification [HIVE-20703](#), qui soumet l'optimisation des partitions de tri dynamique à une décision basée sur les coûts. Si votre charge de travail ne nécessite pas le tri des partitions dynamiques, nous vous recommandons de définir la propriété `hive.optimize.sort.dynamic.partition.threshold` sur `-1` pour désactiver la nouvelle fonctionnalité et obtenir le nombre de réducteurs correctement calculé. Ce problème est résolu dans OSS Hive dans le cadre de la version [HIVE-22269](#) et dans Amazon 6.10.0. EMR
- Hive peut subir une perte de données lorsque vous l'utilisez HDFS comme répertoire temporaire et que vous avez activé la fusion de petits fichiers et que la table contient des chemins de partition statiques.
- Cette version corrige un problème de performance avec Hive si la fusion de petits fichiers (désactivée par défaut) est activée à la fin du ETL travail.
- Cette version corrige un problème de régulation du côté de Glue lorsqu'il n'existe aucune fonction définie par l'utilisateur (`glue.udf`). UDF
- Cette version corrige un problème qui supprime les journaux des conteneurs par le service d'agrégation des journaux des nœuds avant que le lanceur de journaux ne puisse les envoyer vers S3 en cas de YARN mise hors service.
- Cette version corrige la gestion des fichiers compactés/archivés avec un suivi permanent des fichiers de stockage pour. HBase

- Cette version corrige un problème qui affectait les performances de Spark lorsque vous définissez une valeur `true` par défaut pour la configuration `spark.yarn.heterogeneousExecutors.enabled` dans `spark-defaults.conf`.
- Cette version corrige un problème lié à l'échec de la lecture des données de shuffle par Reduce Task. Ce problème provoquait des échecs de requêtes Hive avec une erreur de mémoire corrompue.
- Cette version corrige un problème qui provoquait l'échec du fournisseur de nœuds si le service HDFS NameNode (NN) était bloqué en mode sécurisé lors du remplacement du nœud.
- Cette version ajoute un nouveau mécanisme de nouvelle tentative au flux de travail de dimensionnement des EMR clusters pour les clusters qui exécutent Presto ou Trino. Cette amélioration réduit le risque que le redimensionnement du cluster soit bloqué indéfiniment en raison de l'échec d'une seule opération de redimensionnement. Cela améliore également l'utilisation du cluster, car celui-ci augmente et diminue la capacité plus rapidement.
- Cette version améliore la logique de réduction du cluster afin que votre cluster ne tente pas de réduire les nœuds principaux en dessous du paramètre de facteur de HDFS réplication défini pour le cluster. Cela répond à vos exigences en matière de redondance des données et réduit le risque de blocage d'une opération de dimensionnement.
- Le démon de gestion des journaux a été mis à niveau pour identifier tous les journaux en cours d'utilisation avec des descripteurs de fichiers ouverts sur le stockage d'instance local, ainsi que les processus associés. Cette mise à niveau garantit qu'Amazon supprime EMR correctement les fichiers et récupère de l'espace de stockage une fois les journaux archivés dans Amazon S3.
- Cette version inclut une amélioration du démon de gestion des journaux qui supprime les répertoires d'étapes vides et inutilisés dans le système de fichiers du cluster local. Un trop grand nombre de répertoires vides peut dégrader les performances des EMR démons Amazon et entraîner une surutilisation des disques.
- Cette version résout un problème qui peut survenir lorsque vous créez un nœud périphérique en répliquant l'un des nœuds primaires à partir d'un cluster comportant plusieurs nœuds primaires. Le nœud périphérique répliqué peut retarder les opérations de réduction d'échelle ou entraîner une utilisation élevée de la mémoire sur les nœuds primaires. Pour plus d'informations sur la création d'un nœud périphérique pour communiquer avec votre EMR cluster, consultez la section [Edge Node Creator](#) dans le `aws-samples` référentiel sur GitHub.
- Cette version améliore le processus d'automatisation EMR utilisé par Amazon pour remonter les EBS volumes Amazon sur une instance après un redémarrage.
- Cette version corrige un problème qui entraînait des écarts intermittents dans les métriques Hadoop publiées par Amazon sur EMR Amazon. CloudWatch

- Cette version résout un problème lié aux EMR clusters où une mise à jour du fichier de YARN configuration contenant la liste d'exclusion des nœuds du cluster est interrompue en raison d'une surutilisation du disque. La mise à jour incomplète entrave les futures opérations de réduction de la taille du cluster. Cette version garantit que votre cluster reste sain et que les opérations de dimensionnement fonctionnent comme prévu.
- Cette version améliore le démon de gestion des journaux sur le cluster afin de surveiller des dossiers de journaux supplémentaires dans votre EMR cluster. Cette amélioration permet de minimiser les scénarios de surutilisation des disques.
- Cette version redémarre automatiquement le démon de gestion des journaux sur le cluster lorsqu'il s'arrête. Cette amélioration réduit le risque que les nœuds apparaissent défectueux en raison d'une surutilisation du disque.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Pacifique) (Sydney), Asie-Pacifique (Jakarta) , Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023.0926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (Centre), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (Centre), Israël (Tel Aviv)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (Centre), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

Versions des composants 6.9.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.1.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.6.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.23.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.2.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2,54,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.15.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.15.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.3.3-amzn-1.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.3-amzn-1.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.3-amzn-1.1	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	3.3.3-amzn-1.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.3-amzn-1.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	3.3.3-amzn-1.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.3-amzn-1.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-1.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-1.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.3-amzn-1.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.3-amzn-1.1	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	2.4.13-amzn-0.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.13-amzn-0.1	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.13-amzn-0.1	HBaseclient en ligne de commande.
hbase-rest-server	2.4.13-amzn-0.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.13-amzn-0.1	Service fournissant un point de terminaison Thrift pour HBase
hbase-operator-tools	2.4.13-amzn-0.1	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-2.1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-2.1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-2.1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-2.1	Client de ligne de commande Hive.

Composant	Version	Description
hive-hbase	3.1.3-amzn-2.1	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-2.1	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-2.1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.12.1-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.12.1-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.12.1-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.12.1-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	0.14.1-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.

Composant	Version	Description
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,7.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	6,0.0- SNAPSHOT	Connecteurs Apache Phoenix pour Spark-3

Composant	Version	Description
phoenix-query-server	6.0.0	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.276-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.276-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.276-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	398-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	398-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	398-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.

Composant	Version	Description
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.3.0-amzn-1.1	Clients de ligne de commande Spark.
spark-history-server	3.3.0-amzn-1.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.3.0-amzn-1.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.3.0-amzn-1.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	22.08.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.10.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-0.1	L'YARNapplication Tez et les bibliothèques.
webserver	2.4.41+	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.9.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.9.1

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the Resource Manager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.

Classifications	Description	Actions de reconfiguration
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.

Classifications	Description	Actions de reconfiguration
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.
presto-connector-lakeformat ion	Modifiez les valeurs dans le fichier lakeformation.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-delta	Modification des valeurs dans le fichier delta.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier exchange-manager.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier iceberg.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.

Classifications	Description	Actions de reconfiguration
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélection nés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.

Classifications	Description	Actions de reconfiguration
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.

Classifications	Description	Actions de reconfiguration
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.9.1

Journal des modifications pour la version 6.9.1 et les notes de mise à jour

Date	Événement	Description
30/08/2023	Mise à jour des notes de mise à jour	Ajout de plusieurs corrections liées au plan de contrôle dans les notes de mise à jour
21/08/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 6.9.1 publiées pour la première fois
16/08/2023	Déploiement terminé	Amazon EMR 6.9.1 entièrement déployé dans toutes les régions prises en charge
04/08/2023	Première version	Amazon EMR 6.9.1 a été déployé pour la première fois dans des régions commerciales limitées

Amazon EMR version 6.9.0

Versions d'application 6.9.0

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
Kit AWS SDK pour Java	1,12,170	1,12,170	1,12,170	1,12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,15
AmazonCloudWatchAgent	-	-	-	-
Delta	2.1.0	-	-	-
Flink	1.15.2	1.15.1	1.15.1	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.13	2.4.12	2.4.12	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.3
Hadoop	3.3.3	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.3
Hudi	0.12.1-amzn-0	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.1-amzn-0	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.9.1	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,276	0,273	0,273	0,272
Spark	3.3.0	3.3.0	3.3.0	3.2.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.10.0	2.9.1	2.9.1	2.4.1
Tez	0,1,2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	388	388	378
Zeppelin	0,1,1	0,1,1	0,1,1	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.7

Notes de mise à jour 6.9.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.9.0 d'Amazon. Les modifications concernent la EMR version 6.8.0 d'Amazon. Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Nouvelles fonctions

- La EMR version 6.9.0 d'Amazon prend en charge Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 et Tez 0.10.2.
- La EMR version 6.9.0 d'Amazon inclut une nouvelle application open source, [Delta Lake](#) la 2.1.0.
- L'intégration Amazon Redshift pour Apache Spark est incluse dans les EMR versions 6.9.0 et ultérieures d'Amazon. Auparavant un outil open-source, l'intégration native est un connecteur Spark que vous pouvez utiliser pour créer des applications Apache Spark capables de lire et d'écrire des données sur Amazon Redshift et Amazon Redshift sans serveur. Pour plus d'informations, consultez [Utilisation de l'intégration d'Amazon Redshift pour Apache Spark avec Amazon EMR](#).
- La EMR version 6.9.0 d'Amazon ajoute la prise en charge de l'archivage des journaux dans Amazon S3 lors de la réduction du cluster. Auparavant, vous pouviez uniquement archiver les fichiers journaux sur Amazon S3 lors de la résiliation du cluster. Cette nouvelle fonctionnalité garantit que les fichiers journaux générés sur le cluster sont conservés sur Amazon S3 même après la résiliation du nœud. Pour plus d'informations, consultez [Configuration de la journalisation et du débogage de cluster](#).
- Pour prendre en charge les requêtes de longue durée, Trino inclut désormais un mécanisme d'exécution tolérant aux pannes. L'exécution tolérante aux pannes atténue les échecs des requêtes en réessayant les requêtes qui ont échoué ou les tâches correspondantes. Pour plus d'informations, consultez [Exécution tolérante aux pannes dans Trino](#).
- Vous pouvez utiliser Apache Flink sur Amazon EMR pour unifier BATCH et STREAM traiter les tables Apache Hive ou les métadonnées de n'importe quelle source de table Flink telle qu'Iceberg, Kinesis ou Kafka. Vous pouvez spécifier le catalogue de données AWS Glue comme métastore pour Flink à l'aide de AWS Management Console, AWS CLI ou Amazon EMR API. Pour plus d'informations, consultez [Configuration de Flink sur Amazon EMR](#).
- Vous pouvez désormais spécifier des rôles d'exécution AWS Identity and Access Management (IAM) et un contrôle d'accès AWS Lake Formation basé pour les requêtes Apache Spark, Apache Hive et Presto EMR sur Amazon sur des EC2 clusters avec Amazon SageMaker Studio. Pour plus d'informations, consultez [Configurer les rôles d'exécution pour Amazon EMR Steps](#).

Problèmes connus

- Pour la EMR version 6.9.0 d'Amazon, Trino ne fonctionne pas sur les clusters activés pour Apache Ranger. Si vous devez utiliser Trino avec Ranger, contactez [AWS Support](#).
- Si vous utilisez l'intégration Amazon Redshift à Apache Spark et que vous disposez d'un champ de type heure, timetz, horodatage ou timestamptz avec une précision de l'ordre de la microseconde au format Parquet, le connecteur arrondit les valeurs temporelles à la milliseconde la plus proche. Pour contourner le problème, utilisez le paramètre `unload_s3_format` de format de téléchargement du texte.
- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

- Les connexions aux EMR clusters Amazon depuis Amazon SageMaker Studio peuvent échouer par intermittence avec un code de réponse 403 Forbidden. Cette erreur se produit lorsque la configuration du IAM rôle sur le cluster prend plus de 60 secondes. Pour contourner le problème, vous pouvez installer un EMR correctif Amazon pour activer les nouvelles tentatives et augmenter le délai d'expiration à un minimum de 300 secondes. Suivez les étapes ci-dessous pour appliquer l'action d'amorçage lorsque vous lancez votre cluster.
 1. Téléchargez le script et RPM les fichiers bootstrap depuis l'Amazon S3 URIs suivant.

```
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/replace-rpms.sh
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/emr-secret-agent-1.18.0-SNAPSHOT20221121212949.noarch.rpm
```

2. Chargez les fichiers de l'étape précédente sur un compartiment Amazon S3 dont vous êtes propriétaire. Le compartiment doit se trouver à l' Région AWS endroit où vous prévoyez de lancer le cluster.
3. Incluez l'action d'amorçage suivante lorsque vous lancez votre EMR cluster. Remplacez *bootstrap_URI* and *RPM_URI* avec le correspondant URIs provenant d'Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- Avec les EMR versions 5.36.0 et 6.6.0 à 6.9.0 d'Amazon, SecretAgent les composants du RecordServer service peuvent subir une perte de données de journal en raison d'une configuration incorrecte du modèle de nom de fichier dans les propriétés Log4j2. En cas de configuration incorrecte, les composants ne génèrent qu'un seul fichier journal par jour. Lorsque la stratégie de rotation est appliquée, elle remplace le fichier existant au lieu de générer un nouveau fichier journal comme prévu. Pour contourner le problème, utilisez une action d'amorçage pour générer des journaux toutes les heures et ajoutez un nombre entier auto-incrémenté dans le nom du fichier pour gérer la rotation.

Pour les versions Amazon EMR 6.6.0 à 6.9.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Pour Amazon EMR 5.36.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Apache Flink fournit des FileSystem connecteurs natifs S3 FileSystem et Hadoop, qui permettent aux applications de créer FileSink et d'écrire les données dans Amazon S3. Cela FileSink échoue avec l'une des deux exceptions suivantes.

```
java.lang.UnsupportedOperationException: Recoverable writers on Hadoop are only supported for HDFS
```

```
Caused by: java.lang.NoSuchMethodError:
  org.apache.hadoop.io.retry.RetryPolicies.retryOtherThanRemoteAndSaslException(Lorg/
  apache/hadoop/io/retry/RetryPolicy;Ljava/util/Map;)Lorg/apache/hadoop/io/retry/
  RetryPolicy;
                                at
  org.apache.hadoop.yarn.client.RMProxy.createRetryPolicy(RMProxy.java:302) ~[hadoop-
  yarn-common-3.3.3-amzn-0.jar:?]
```

Pour contourner le problème, vous pouvez installer un EMR correctif Amazon, qui corrige le problème ci-dessus dans Flink. Suivez les étapes suivantes pour appliquer l'action d'amorçage lors du lancement de votre cluster.

1. Téléchargez le fichier flink-rpm dans votre compartiment Amazon S3. Votre RPM voie l'est `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Téléchargez le script et les RPM fichiers bootstrap depuis Amazon S3 en utilisant ce qui suit URL. Remplacez `regionName` par l' Région AWS endroit où vous prévoyez de lancer le cluster.

```
s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-rpms.sh
```

3. Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Dans Amazon EMR 6.8.0 et 6.9.0, cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans le dimensionnement géré activé.

Avec [Amazon EMR 6.10.0](#), il existe une solution à ce problème qui consiste à définir la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications to in. false` `yarn-site.xml` Dans les EMR versions 6.11.0 et supérieures d'Amazon ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la configuration est définie `false` par défaut pour résoudre ce problème.

Modifications, améliorations et problèmes résolus

- Pour les EMR versions 6.9.0 et ultérieures d'Amazon, tous les composants installés par Amazon EMR qui utilisent les bibliothèques Log4j utilisent Log4j version 2.17.1 ou ultérieure.
- Lorsque vous utilisez le connecteur DynamoDB avec Spark on EMR Amazon versions 6.6.0, 6.7.0 et 6.8.0, toutes les lectures de votre table renvoient un résultat vide, même si le split d'entrée fait référence à des données non vides. La EMR version 6.9.0 d'Amazon résout ce problème.
- Amazon EMR 6.9.0 ajoute une prise en charge limitée du contrôle d'accès basé sur Lake Formation avec Apache Hudi lors de la lecture de données à l'aide de Spark. SQL Le support concerne les SELECT requêtes utilisant Spark SQL et se limite au contrôle d'accès au niveau des colonnes. Pour plus d'informations, consultez [Hudi et Lake Formation](#).
- Lorsque vous utilisez Amazon EMR 6.9.0 pour créer un cluster Hadoop avec les [étiquettes de nœuds](#) activées, les [YARNmétriques API](#) renvoient des informations agrégées sur toutes les partitions, au lieu de la partition par défaut. Pour plus d'informations, consultez [YARN-11414](#).
- Avec la EMR version 6.9.0 d'Amazon, nous avons mis à jour Trino vers la version 398, qui utilise Java 17. La précédente version prise en charge de Trino pour Amazon EMR 6.8.0 était Trino 388 fonctionnant sous Java 11. Pour plus d'informations sur cette modification, consultez [Mises à jour de Trino vers Java 17](#) sur le blog de Trino.
- Cette version corrige un problème d'inadéquation des séquences temporelles entre Apache BigTop et Amazon EMR lors de la séquence de démarrage EC2 du cluster. Ce décalage se produit lorsqu'un système tente d'effectuer deux ou plusieurs opérations en même temps au lieu de les effectuer dans le bon ordre. Par conséquent, certaines configurations de cluster ont connu des délais de démarrage des instances et des temps de démarrage des clusters plus lents.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

Note

Cette version ne reçoit plus de AMI mises à jour automatiques puisqu'elle a été remplacée par une autre version de correctif. La version du correctif est indiquée par le numéro qui suit la deuxième décimale (6.8.1). Pour savoir si vous utilisez la dernière version du correctif, consultez les versions disponibles dans le [guide des versions](#), ou consultez le menu déroulant des EMRversions d'Amazon lorsque vous créez un cluster dans la console, ou utilisez l'[list-release-labels](#) CLI action [ListReleaseLabels](#) API.

Pour être tenu au courant des nouvelles versions, abonnez-vous au RSS fil d'actualité sur le [site Quoi de neuf ?](#) page.

OsRelea Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre), Israël (Tel Aviv)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 307.0	4,14,305	15 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 207.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202210.1	4,14,301	12 janvier 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 103.3	4,14.296	5 décembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

Versions des composants 6.9.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.1.0	Delta Lake est un format de table ouvert pour les jeux de données analytiques de grande taille.
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.6.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.23.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.2.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2,54,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.15.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.15.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.3.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.3.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.3.3-amzn-1	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	3.3.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.3.3-amzn-1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	3.3.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.3.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.3.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.3.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.3.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	2.4.13-amzn-0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.13-amzn-0	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.4.13-amzn-0	HBaseclient en ligne de commande.
hbase-rest-server	2.4.13-amzn-0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.13-amzn-0	Service fournissant un point de terminaison Thrift pour HBase
hbase-operator-tools	2.4.13-amzn-0	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-2	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-2	Client de ligne de commande Hive.

Composant	Version	Description
hive-hbase	3.1.3-amzn-2	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.12.1-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.12.1-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.12.1-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.12.1-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	0.14.1-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.

Composant	Version	Description
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,7.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	6,0.0- SNAPSHOT	Connecteurs Apache Phoenix pour Spark-3

Composant	Version	Description
phoenix-query-server	6.0.0	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.276-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.276-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.276-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	398-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	398-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	398-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.

Composant	Version	Description
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.3.0-amzn-1	Clients de ligne de commande Spark.
spark-history-server	3.3.0-amzn-1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.3.0-amzn-1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.3.0-amzn-1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	22.08.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.10.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.10.2-amzn-0	L'YARNapplication Tez et les bibliothèques.
webserver	2.4.41+	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.9.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.9.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.

Classifications	Description	Actions de reconfiguration
delta-defaults	Modification des valeurs dans le fichier delta-defaults.conf de Delta.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j2	Modifiez les paramètres Livy log4j2.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.

Classifications	Description	Actions de reconfiguration
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.
presto-connector-lakeformation	Modifiez les valeurs dans le fichier lakeformation.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-delta	Modification des valeurs dans le fichier delta.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Modification des valeurs dans le fichier exchange-manager.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier iceberg.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.

Classifications	Description	Actions de reconfiguration
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélection nés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.

Classifications	Description	Actions de reconfiguration
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.

Classifications	Description	Actions de reconfiguration
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.9.0

Journal des modifications pour la version 6.9.0 et les notes de mise à jour

Date	Événement	Description
30/08/2023	Mise à jour des notes de mise à jour	Ajout d'un correctif pour un problème de non-concordance des séquences de synchronisation
21/08/2023	Mise à jour des notes de mise à jour	Ajout d'un problème connu avec Hadoop 3.3.3.
26/07/2023	Mettre à jour	Étiquettes 2.0.20230612.0 et 2.0.20230628.0 de la nouvelle version du système d'exploitation.
13/12	Mise à jour des notes de mise à jour	Fonctionnalité ajoutée et problème connu d'exécution avec SageMaker
29/11	Mise à jour des notes de mise à jour et de la documentation	Ajout d'une fonctionnalité pour l'intégration d'Amazon Redshift pour Apache Spark
23/11	Mise à jour des notes de mise à jour	Suppression de l'entrée Log4j

Date	Événement	Description
11/18	Déploiement terminé	Amazon EMR 6.9 entièrement déployé dans toutes les régions prises en charge
11/18	Publication de documents	Notes de mise à jour d'Amazon EMR 6.9 publiées pour la première fois
14/11	Première version	Amazon EMR 6.9 déployé dans un nombre limité de régions commerciales

Amazon EMR version 6.8.1

Versions d'application 6.8.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
Kit AWS SDK pour Java	1,12,170	1,12,170	1,12,170	1,12,170
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,15	2,1,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.15.1	1.15.1	1.14.2	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.12	2.4.12	2.4.4	2.4.4
HCatalog	3.1.3	3.1.3	3.1.3	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.3	3.1.2
Hudi	0.11.1-amzn-0	0.11.1-amzn-0	0.11.0-amzn-0	0.10.1-amzn-0
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	0.14.0-amzn-0	0.14.0-amzn-0	0.13.1-amzn-0	0.13.1
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.9.1	1.8.0	1.8.0

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,273	0,273	0,272	0,267
Spark	3.3.0	3.3.0	3.2.1	3.2.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.9.1	2.9.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	388	388	378	367
Zeppelin	0,1,1	0,1,1	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.7	3.5.7

Notes de mise à jour 6.8.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.8.1 d'Amazon. Les modifications ont été apportées à la version 6.8.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.8.1](#).

Modifications, améliorations et problèmes résolus

- Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche.

Cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans activation de la mise à l'échelle gérée.

Dans les EMR versions 6.11.0 et supérieures d'Amazon ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` est définie sur `false` in `yarn-site.xml` pour résoudre ce problème.

Bien que le correctif corrige les problèmes introduits par YARN -9608, il peut entraîner l'échec des tâches Hive en raison de la perte de données par shuffle sur les clusters sur lesquels le dimensionnement géré est activé. Nous avons atténué ce risque dans cette version en définissant également `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` pour les charges de travail Hive. Cette configuration n'est disponible qu'avec les EMR versions 6.11.0 et supérieures d'Amazon.

- Le collecteur de métriques n'enverra aucune métrique au plan de contrôle après le basculement du nœud primaire dans les clusters avec la configuration des groupes d'instances.
- Cette version élimine les tentatives en cas d'échec des HTTP demandes adressées aux points de terminaison du collecteur de mesures.
- Cette version inclut une modification qui permet aux clusters à haute disponibilité de se remettre d'un état défaillant après le redémarrage.
- Cette version corrige un problème selon lequel de grandes quantités créées par l'utilisateur UIDs provoquaient des exceptions de dépassement.
- Cette version corrige les problèmes de temporisation liés au processus de EMR reconfiguration d'Amazon.
- Cette version permet d'éviter un problème selon lequel l'échec de la reconfiguration pourrait interrompre d'autres processus indépendants.
- Cette version inclut des correctifs de sécurité.
- Cette version résout un problème selon lequel les clusters exécutant des charges de travail sur Spark avec Amazon EMR peuvent recevoir silencieusement des résultats incorrects avec `contains`, `startsWith`, `endsWith`, `et. like`. Ce problème se produit lorsque vous utilisez les expressions sur des champs partitionnés contenant des métadonnées sur le serveur Amazon EMR Hive3 Metastore (`. HMS`).
- Avec Amazon EMR 6.6.0 à 6.9.x, les INSERT requêtes comportant une partition dynamique et une clause ORDER BY ou SORT BY auront toujours deux réducteurs. Ce problème est dû à la OSS modification [HIVE-20703](#), qui soumet l'optimisation des partitions de tri dynamique à une décision basée sur les coûts. Si votre charge de travail ne nécessite

pas le tri des partitions dynamiques, nous vous recommandons de définir la propriété `hive.optimize.sort.dynamic.partition.threshold` sur `-1` pour désactiver la nouvelle fonctionnalité et obtenir le nombre de réducteurs correctement calculé. Ce problème est résolu dans OSS Hive dans le cadre de la version [HIVE-22269](#) et dans Amazon 6.10.0. EMR

- Hive peut subir une perte de données lorsque vous l'utilisez HDFS comme répertoire temporaire et que vous avez activé la fusion de petits fichiers et que la table contient des chemins de partition statiques.
- Cette version corrige un problème de performance avec Hive si la fusion de petits fichiers (désactivée par défaut) est activée à la fin du ETL travail.
- Cette version corrige un problème de régulation du côté de Glue lorsqu'il n'existe aucune fonction définie par l'utilisateur (). UDF
- Cette version corrige un problème qui supprime les journaux des conteneurs par le service d'agrégation des journaux des nœuds avant que le lanceur de journaux ne puisse les envoyer vers S3 en cas de YARN mise hors service.
- Cette version corrige la gestion des fichiers compactés/archivés avec un suivi permanent des fichiers de stockage pour. HBase
- Cette version corrige un problème qui affectait les performances de Spark lorsque vous définissez une valeur `true` par défaut pour la configuration `spark.yarn.heterogeneousExecutors.enabled` dans `spark-defaults.conf`.
- Cette version corrige un problème lié à l'échec de la lecture des données de shuffle par Reduce Task. Ce problème provoquait des échecs de requêtes Hive avec une erreur de mémoire corrompue.
- Cette version corrige un problème qui provoquait l'échec du fournisseur de nœuds si le service HDFS NameNode (NN) était bloqué en mode sécurisé lors du remplacement du nœud.
- Cette version ajoute un nouveau mécanisme de nouvelle tentative au flux de travail de dimensionnement des EMR clusters pour les clusters qui exécutent Presto ou Trino. Cette amélioration réduit le risque que le redimensionnement du cluster soit bloqué indéfiniment en raison de l'échec d'une seule opération de redimensionnement. Cela améliore également l'utilisation du cluster, car celui-ci augmente et diminue la capacité plus rapidement.
- Cette version améliore la logique de réduction du cluster afin que votre cluster ne tente pas de réduire les nœuds principaux en dessous du paramètre de facteur de HDFS réplication défini pour le cluster. Cela répond à vos exigences en matière de redondance des données et réduit le risque de blocage d'une opération de dimensionnement.

- Le démon de gestion des journaux a été mis à niveau pour identifier tous les journaux en cours d'utilisation avec des descripteurs de fichiers ouverts sur le stockage d'instance local, ainsi que les processus associés. Cette mise à niveau garantit qu'Amazon supprime EMR correctement les fichiers et récupère de l'espace de stockage une fois les journaux archivés dans Amazon S3.
- Cette version inclut une amélioration du démon de gestion des journaux qui supprime les répertoires d'étapes vides et inutilisés dans le système de fichiers du cluster local. Un trop grand nombre de répertoires vides peut dégrader les performances des EMR démons Amazon et entraîner une surutilisation des disques.
- Cette version résout un problème qui peut survenir lorsque vous créez un nœud périphérique en répliquant l'un des nœuds primaires à partir d'un cluster comportant plusieurs nœuds primaires. Le nœud périphérique répliqué peut retarder les opérations de réduction d'échelle ou entraîner une utilisation élevée de la mémoire sur les nœuds primaires. Pour plus d'informations sur la création d'un nœud périphérique pour communiquer avec votre EMR cluster, consultez la section [Edge Node Creator](#) dans le `aws-samples` référentiel sur GitHub.
- Cette version améliore le processus d'automatisation EMR utilisé par Amazon pour remonter les EBS volumes Amazon sur une instance après un redémarrage.
- Cette version corrige un problème qui entraînait des écarts intermittents dans les métriques Hadoop publiées par Amazon sur EMR Amazon. CloudWatch
- Cette version résout un problème lié aux EMR clusters où une mise à jour du fichier de YARN configuration contenant la liste d'exclusion des nœuds du cluster est interrompue en raison d'une surutilisation du disque. La mise à jour incomplète entrave les futures opérations de réduction de la taille du cluster. Cette version garantit que votre cluster reste sain et que les opérations de dimensionnement fonctionnent comme prévu.
- Cette version améliore le démon de gestion des journaux sur le cluster afin de surveiller des dossiers de journaux supplémentaires dans votre EMR cluster. Cette amélioration permet de minimiser les scénarios de surutilisation des disques.
- Cette version redémarre automatiquement le démon de gestion des journaux sur le cluster lorsqu'il s'arrête. Cette amélioration réduit le risque que les nœuds apparaissent défectueux en raison d'une surutilisation du disque.
- Cette version ajoute la prise en charge de l'archivage des journaux sur Amazon S3 lors de la réduction de la capacité du cluster. Auparavant, vous pouviez uniquement archiver les fichiers journaux sur Amazon S3 lors de la résiliation du cluster. Cette nouvelle fonctionnalité garantit que les fichiers journaux générés sur le cluster sont conservés sur Amazon S3 même après la résiliation du nœud. Pour plus d'informations, consultez [Configuration de la journalisation et du débogage de cluster](#).

- Cette version corrige un problème qui s'est produit lorsque l'Amazon S3 URI pour une action de démarrage s'est terminé par un numéro de port, par exemple :a . b . c . d : 4345. Amazon EMR les analysait incorrectementURIs, de sorte que toutes les actions d'amorçage associées échoueraient.
- Cette version corrige un problème d'inadéquation des séquences temporelles entre Apache BigTop et Amazon EMR lors de la séquence de démarrage EC2 du cluster. Ce décalage se produit lorsqu'un système tente d'effectuer deux ou plusieurs opérations en même temps au lieu de les effectuer dans le bon ordre. Par conséquent, certaines configurations de cluster ont connu des délais de démarrage des instances et des temps de démarrage des clusters plus lents.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap),

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			<p>Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv),</p>

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (centre), Israël (Tel Aviv), (États-Unis ouest), AWS GovCloud AWS GovCloud (États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne) Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne) Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne) Afrique (Le Cap), Amérique du Sud (São Paulo), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

Versions des composants 6.8.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.22.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.1.0	EMRConnecteur S3Select
emrfs	2,53,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
flink-client	1.15.1	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.15.1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-8.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-8.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-8.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-8.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-hdfs-journalnode	3.2.1-amzn-8.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httfs-server	3.2.1-amzn-8.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-8.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-8.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-8.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.2.1-amzn-8.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-8.1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.12-amzn-0.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.12-amzn-0.1	Service pour desservir une ou plusieurs HBase régions.

Composant	Version	Description
hbase-client	2.4.12-amzn-0.1	HBaseclient en ligne de commande.
hbase-rest-server	2.4.12-amzn-0.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.12-amzn-0.1	Service fournissant un point de terminaison Thrift pour. HBase
hbase-operator-tools	2.4.12-amzn-0.1	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-1.1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-1.1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-1.1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-1.1	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-1.1	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-1.1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server2	3.1.3-amzn-1.1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.11.1-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.11.1-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.11.1-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.11.1-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	0.14.0-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark

Composant	Version	Description
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,7.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.2	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.2	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.273.3-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.273.3-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.273.3-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	388-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	388-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	388-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger

Composant	Version	Description
spark-client	3.3.0-amzn-0.1	Clients de ligne de commande Spark.
spark-history-server	3.3.0-amzn-0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.3.0-amzn-0.1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.3.0-amzn-0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	22.06.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.9.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.8.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.8.1

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the Resource Manager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classifications	Description	Actions de reconfiguration
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.

Classifications	Description	Actions de reconfiguration
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.

Classifications	Description	Actions de reconfiguration
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier <code>iceberg.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.

Classifications	Description	Actions de reconfiguration
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.

Classifications	Description	Actions de reconfiguration
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.8.1

Journal des modifications pour la version 6.8.1 et les notes de mise à jour

Date	Événement	Description
30/08/2023	Mise à jour des notes de mise à jour	Ajout de plusieurs corrections liées au plan de contrôle dans les notes de mise à jour
21/08/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 6.8.1 publiées pour la première fois
16/08/2023	Déploiement terminé	Amazon EMR 6.8.1 entièrement déployé dans toutes les régions prises en charge
04/08/2023	Première version	Amazon EMR 6.8.1 a été déployé pour la première fois dans des régions commerciales limitées

Amazon EMR version 6.8.0

Versions d'application 6.8.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)

- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
Kit AWS SDK pour Java	1,12,170	1,12,170	1,12,170	1,12,31
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,15	2,1,10	2,1,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.15.1	1.14.2	1.14.2	1.14.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.12	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.3	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.3	3.1.2	3.1.2
Hudi	0.11.1-amzn-0	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1
Hue	4.10.0	4.10.0	4.10.0	4.9.0
Iceberg	0.14.0-amzn-0	0.13.1-amzn-0	0.13.1	0.12.0
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.9.1	1.8.0	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,273	0,272	0,267	0,261
Spark	3.3.0	3.2.1	3.2.0	3.1.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.9.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	388	378	367	360
Zeppelin	0,1,1	0.10.0	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.7	3.5.7	3.5.7

Notes de mise à jour 6.8.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.8.0 d'Amazon. Les modifications ont été apportées à la version 6.7.0.

Nouvelles fonctions

- La fonctionnalité Amazon EMR Steps prend désormais en charge le point de terminaison JDBC et/ou les ODBC clients Apache Livy. Pour plus d'informations, consultez [Configurer les rôles d'exécution pour Amazon EMR Steps](#).

- La EMR version 6.8.0 d'Amazon est fournie avec la HBase version 2.4.12 d'Apache. Avec cette HBase version, vous pouvez à la fois archiver et supprimer vos HBase tables. Le processus d'archivage Amazon S3 renomme tous les fichiers de table dans le répertoire d'archive. Ce processus peut être long et coûteux. Vous pouvez désormais ignorer le processus d'archivage et supprimer rapidement des tables volumineuses. Pour plus d'informations, consultez [Utilisation de la HBase coque](#).

Problèmes connus

- Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Dans Amazon EMR 6.8.0 et 6.9.0, cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans le dimensionnement géré activé.

Avec [Amazon EMR 6.10.0](#), il existe une solution à ce problème qui consiste à définir la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `in.false` dans `yarn-site.xml`. Dans les EMR versions 6.11.0 et supérieures d'Amazon ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la configuration est définie `false` par défaut pour résoudre ce problème.

Modifications, améliorations et problèmes résolus

- Lorsque les EMR versions 6.5.0, 6.6.0 ou 6.7.0 d'Amazon lisaient les tables Apache Phoenix via le shell Apache Spark, Amazon EMR a produit un `NoSuchMethodError`. La EMR version 6.8.0 d'Amazon résout ce problème.
- La EMR version 6.8.0 d'Amazon est fournie avec [Apache Hudi](#) 0.11.1 ; toutefois, les clusters Amazon EMR 6.8.0 sont également compatibles avec l'open source de Hudi 0.12.0. `hudi-spark3.3-bundle_2.12`
- La EMR version 6.8.0 d'Amazon est fournie avec Apache Spark 3.3.0. Cette version de Spark utilise Apache Log4j 2 et le fichier `log4j2.properties` pour configurer Log4j dans les processus Spark. Si vous utilisez Spark dans le cluster ou si vous créez des EMR clusters avec des paramètres de configuration personnalisés, et que vous souhaitez passer à la EMR version 6.8.0 d'Amazon, vous devez migrer vers la nouvelle classification de `spark-log4j2` configuration

et le nouveau format de clé pour Apache Log4j 2. Pour plus d'informations, consultez [Migration d'Apache Log4j 1.x vers Log4j 2.x](#).

- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

Note

Cette version ne reçoit plus de AMI mises à jour automatiques puisqu'elle a été remplacée par une autre version de correctif. La version du correctif est indiquée par le numéro qui suit la deuxième décimale (6 . 8 . **1**). Pour savoir si vous utilisez la dernière version du correctif, consultez les versions disponibles dans le [guide des versions](#), ou consultez le menu déroulant des EMRversions d'Amazon lorsque vous créez un cluster dans la console, ou utilisez l'[list-release-labels](#) CLI action [ListReleaseLabels](#) API ou. Pour être tenu au courant des nouvelles versions, abonnez-vous au RSS fil d'actualité sur le [site Quoi de neuf ?](#) page.

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Paci

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			<p>Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie-Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie-Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv)</p>

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202304.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 307.0	4,14,305	15 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 207.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023119.1	4,14,301	3 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202210.1	4,14,301	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 103.3	4,14.296	5 décembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 004.0	4,14.294	2 novembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 912.1	4,14,291	6 septembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

Problèmes connus

- Lorsque vous utilisez le connecteur DynamoDB avec Spark on EMR Amazon versions 6.6.0, 6.7.0 et 6.8.0, toutes les lectures de votre table renvoient un résultat vide, même si le split d'entrée fait référence à des données non vides. Cela est dû au fait que Spark 3.2.0 définit `spark.hadoopRDD.ignoreEmptySplits` sur `true` par défaut. Pour contourner le problème,

définissez explicitement `spark.hadoopRDD.ignoreEmptySplits` sur `false`. La EMR version 6.9.0 d'Amazon résout ce problème.

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

- Avec les EMR versions 5.36.0 et 6.6.0 à 6.9.0 d'Amazon, `SecretAgent` les composants du `RecordServer` service peuvent subir une perte de données de journal en raison d'une configuration incorrecte du modèle de nom de fichier dans les propriétés `Log4j2`. En cas de configuration incorrecte, les composants ne génèrent qu'un seul fichier journal par jour. Lorsque la stratégie de rotation est appliquée, elle remplace le fichier existant au lieu de générer un nouveau fichier journal comme prévu. Pour contourner le problème, utilisez une action d'amorçage pour générer des journaux toutes les heures et ajoutez un nombre entier auto-incrémenté dans le nom du fichier pour gérer la rotation.

Pour les versions Amazon EMR 6.6.0 à 6.9.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Pour Amazon EMR 5.36.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Versions des composants 6.8.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	3.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-notebook-env	1.7.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.22.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.1.0	EMRConnecteur S3Select
emrfs	2,53,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.15.1	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.15.1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	3.2.1-amzn-8	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-8	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-8	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-8	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-8	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.2.1-amzn-8	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-8	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-8	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-8	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	3.2.1-amzn-8	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-8	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.12-amzn-0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.12-amzn-0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.4.12-amzn-0	HBaseclient en ligne de commande.
hbase-rest-server	2.4.12-amzn-0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.12-amzn-0	Service fournissant un point de terminaison Thrift pour. HBase
hbase-operator-tools	2.4.12-amzn-0	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.3-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	3.1.3-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-1	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-1	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-1	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.11.1-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.11.1-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.11.1-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.

Composant	Version	Description
hudi-spark	0.11.1-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	0.14.0-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.9.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,7.0	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.2	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.2	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.273.3-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.273.3-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.273.3-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	388-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.

Composant	Version	Description
trino-worker	388-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	388-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.3.0-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.3.0-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.3.0-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.3.0-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	22.06.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.

Composant	Version	Description
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.9.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0,1,1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.10	ZooKeeper client en ligne de commande.

Classifications des configurations 6.8.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration

EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.8.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Actions de reconfiguration
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.

Classifications	Description	Actions de reconfiguration
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.

Classifications	Description	Actions de reconfiguration
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-tpcds	Modifiez les valeurs dans le fichier <code>tpcds.properties</code> de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier <code>trino-env.sh</code> de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier <code>iceberg.properties</code> de Trino.	Restarts Trino-Server (for Trino)

Classifications	Description	Actions de reconfiguration
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.

Classifications	Description	Actions de reconfiguration
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j2	Modifiez les valeurs dans le fichier log4j2.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.

Classifications	Description	Actions de reconfiguration
yarn-env	Modifiez les valeurs dans l'YARN environnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 6.8.0

Journal des modifications pour la version 6.8.0 et les notes de mise à jour

Date	Événement	Description
21/08/2023	Mettre à jour	Ajout d'un problème connu avec Hadoop 3.3.3.
26/07/2023	Mettre à jour	Étiquettes 2.0.20230612.0 et 2.0.20230628.0 de la nouvelle version du système d'exploitation.
05.09-06	Déploiement terminé	Amazon EMR 6.8 entièrement déployé dans toutes les régions prises en charge
05.09-06	Publication initiale	Notes de mise à jour d'Amazon EMR 6.8 publiées pour la première fois
08-31	Première version	Amazon EMR 6.8 est disponible dans un nombre limité de régions commerciales

Amazon EMR version 6.7.0

Versions d'application 6.7.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
Kit AWS SDK pour Java	1,12,170	1,12,170	1,12,31	1,12,31
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,15	2,1,10	2,1,10	2,1,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.0	1.13.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.4.4	2.4.4
HCatalog	3.1.3	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.3	3.1.2	3.1.2	3.1.2
Hudi	0.11.0-amzn-0	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	0.13.1-amzn-0	0.13.1	0.12.0	-

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.1.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,272	0,267	0,261	0,254,1
Spark	3.2.1	3.2.0	3.1.2	3.1.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	378	367	360	359
Zeppelin	0.10.0	0.10.0	0.10.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.5.7	3.5.7

Notes de mise à jour 6.7.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.7.0 d'Amazon. Les modifications ont été apportées à la version 6.6.0.

Date de parution initiale : 15 juillet 2022

Nouvelles fonctions

- Amazon prend EMR désormais en charge Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 et Trino 0.378.
- Prend en charge les contrôles d'accès basés sur les IAM rôles et les formations lacustres avec des EMR étapes (Spark, Hive) pour Amazon EMR sur les clusters. EC2
- Prend en charge les instructions de définition de données Apache Spark sur les clusters compatibles avec Apache Ranger. Cela inclut désormais la prise en charge des applications Trino lisant et écrivant des métadonnées Apache Hive sur des clusters compatibles avec Apache Ranger. Pour plus d'informations, consultez [Activer la gouvernance fédérée à l'aide de Trino et Apache Ranger sur Amazon](#). EMR
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapou

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			r), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 307.0	4,14,305	15 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023-07.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023119.1	4,14,301	3 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202210.1	4,14,301	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 103.3	4,14.296	5 décembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022004.0	4,14.294	2 novembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2022 912.1	4,14,291	7 octobre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0,2022 719,0	4,14,287	10 août 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 606.1	4,14,281	15 juillet 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

Problèmes connus

- Lorsque les EMR versions 6.5.0, 6.6.0 ou 6.7.0 d'Amazon lisent les tables Apache Phoenix via le shell Apache Spark, cela se produit `NoSuchMethodError` car Amazon EMR utilise une valeur incorrecte. `Hbase.compat.version` La EMR version 6.8.0 d'Amazon résout ce problème.
- Lorsque vous utilisez le connecteur DynamoDB avec Spark on EMR Amazon versions 6.6.0, 6.7.0 et 6.8.0, toutes les lectures de votre table renvoient un résultat vide, même si le split

d'entrée fait référence à des données non vides. Cela est dû au fait que Spark 3.2.0 définit `spark.hadoopRDD.ignoreEmptySplits` sur `true` par défaut. Pour contourner le problème, définissez explicitement `spark.hadoopRDD.ignoreEmptySplits` sur `false`. La EMR version 6.9.0 d'Amazon résout ce problème.

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

- Avec les EMR versions 5.36.0 et 6.6.0 à 6.9.0 d'Amazon, `SecretAgent` les composants du `RecordServer` service peuvent subir une perte de données de journal en raison d'une configuration incorrecte du modèle de nom de fichier dans les propriétés `Log4j2`. En cas de configuration incorrecte, les composants ne génèrent qu'un seul fichier journal par jour. Lorsque la stratégie de rotation est appliquée, elle remplace le fichier existant au lieu de générer un nouveau fichier journal comme prévu. Pour contourner le problème, utilisez une action d'amorçage pour générer des journaux toutes les heures et ajoutez un nombre entier auto-incrémenté dans le nom du fichier pour gérer la rotation.

Pour les versions Amazon EMR 6.6.0 à 6.9.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```


Pour Amazon EMR 5.36.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Il `GetClusterSessionCredentials` API n'est pas pris en charge avec les clusters qui s'exécutent sur Amazon EMR 6.7 ou une version antérieure.

Versions des composants 6.7.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	3.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.6.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.22.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.1.0	EMRConnecteur S3Select
emrfs	2,52,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.14.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.14.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-7	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-7	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-7	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-7	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-7	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.2.1-amzn-7	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-7	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-7	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	3.2.1-amzn-7	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.2.1-amzn-7	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-7	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.4-amzn-3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.4-amzn-3	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.4.4-amzn-3	HBaseclient en ligne de commande.
hbase-rest-server	2.4.4-amzn-3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.4-amzn-3	Service fournissant un point de terminaison Thrift pour HBase
hbase-operator-tools	2.4.4-amzn-3	Outil de réparation pour les HBase clusters Apache.

Composant	Version	Description
hcatalog-client	3.1.3-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.3-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.3-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.3-amzn-0	Client de ligne de commande Hive.
hive-hbase	3.1.3-amzn-0	Client Hive-hbase.
hive-metastore-server	3.1.3-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.3-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.11.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.11.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.

Composant	Version	Description
hudi-trino	0.11.0-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.11.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	0.13.1-amzn-0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.8.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,0,194	Pilotes Nvidia et boîte à outils Cuda

Composant	Version	Description
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.2	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.2	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.272-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.272-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.272-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
trino-coordinator	378-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	378-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	378-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.2.1-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.2.1-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.2.1-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.2.1-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
spark-rapids	22.02.0-amzn-1	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARNapplication Tez et les bibliothèques.
webserver	2.4.41+	HTTPserveur Apache.
zeppelin-server	0.10.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.7	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.7	ZooKeeper client en ligne de commande.

Classifications des configurations 6.7.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.7.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.

Classifications	Description	Actions de reconfiguration
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.

Classifications	Description	Actions de reconfiguration
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-tpcds	Modifiez les valeurs dans le fichier <code>tpcds.properties</code> de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier <code>trino-env.sh</code> de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier <code>iceberg.properties</code> de Trino.	Restarts Trino-Server (for Trino)

Classifications	Description	Actions de reconfiguration
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.

Classifications	Description	Actions de reconfiguration
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.

Classifications	Description	Actions de reconfiguration
yarn-env	Modifiez les valeurs dans l'YARN environnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 6.6.0

Versions d'application 6.6.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
Kit AWS SDK pour Java	1,12,170	1,12,31	1,12,31	1,1,977
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.0	1.13.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.4.4	2.2.6
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.10.1-amzn-0	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	0.13.1	0.12.0	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2
Livy	0.7.1	0.7.1	0.7.1	0.7.0
MXNet	1.8.0	1.8.0	1.8.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.1.2	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,267	0,261	0,254,1	0,245,1
Spark	3.2.0	3.1.2	3.1.2	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	367	360	359	350
Zeppelin	0.10.0	0.10.0	0.9.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.5.7	3.4.14

Notes de mise à jour 6.6.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.6.0 d'Amazon. Les modifications ont été apportées à la version 6.5.0.

Date de parution initiale : 9 mai 2022

Date de mise à jour de la documentation : 15 juin 2022

Nouvelles fonctions

- Amazon EMR 6.6 prend désormais en charge Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 et PrestoDB 0.267.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 307.0	4,14,305	15 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023-07.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023119.1	4,14,301	3 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202210.1	4,14,301	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 103.3	4,14.296	5 décembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022004.0	4,14.294	2 novembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2022 912.1	4,14,291	7 octobre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0,2022 805.0	4,14,287	30 août 2022	us-west-1

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2022 719,0	4,14.287	10 août 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 426.0	4,14,281	10 juin 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 406.1	4,14,275	2 mai 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

- [Avec Amazon EMR 6.6 et versions ultérieures, les applications qui utilisent Log4j 1.x et Log4j 2.x sont mises à niveau pour utiliser Log4j 1.2.17 \(ou version ultérieure\) et Log4j 2.17.1 \(ou version ultérieure\) respectivement, et ne nécessitent pas d'utiliser les actions d'amorçage fournies pour atténuer les problèmes.](#) CVE
- [Dimensionnement géré] Optimisation du dimensionnement géré par les données Spark Shuffle - Pour EMR les versions 5.34.0 et ultérieures d'Amazon, ainsi que pour EMR les versions 6.4.0 et ultérieures, le dimensionnement géré prend désormais en compte les données Spark shuffle (données que Spark redistribue entre les partitions pour effectuer des opérations spécifiques).

Pour plus d'informations sur les opérations de shuffle, consultez la section [Utilisation du dimensionnement EMR géré EMR dans Amazon](#) dans le Guide de EMRgestion Amazon et le Guide de [programmation Spark](#).

- À partir d'Amazon EMR 5.32.0 et 6.5.0, le dimensionnement dynamique de l'exécuteur pour Apache Spark est activé par défaut. Pour activer ou désactiver cette fonctionnalité, vous pouvez utiliser le paramètre de configuration `spark.yarn.heterogeneousExecutors.enabled`.

Modifications, améliorations et problèmes résolus

- Amazon EMR réduit le temps de démarrage des clusters de 80 secondes en moyenne pour les clusters qui utilisent l'AMloption EMR par défaut et installent uniquement des applications courantes, telles qu'Apache Hadoop, Apache Spark et Apache Hive.

Problèmes connus

- Lorsque les EMR versions 6.5.0, 6.6.0 ou 6.7.0 d'Amazon lisent les tables Apache Phoenix via le shell Apache Spark, cela se produit `NoSuchMethodError` car Amazon EMR utilise une valeur incorrecte. `Hbase.compat.version` La EMR version 6.8.0 d'Amazon résout ce problème.
- Lorsque vous utilisez le connecteur DynamoDB avec Spark on EMR Amazon versions 6.6.0, 6.7.0 et 6.8.0, toutes les lectures de votre table renvoient un résultat vide, même si le split d'entrée fait référence à des données non vides. Cela est dû au fait que Spark 3.2.0 définit `spark.hadoopRDD.ignoreEmptySplits` sur `true` par défaut. Pour contourner le problème, définissez explicitement `spark.hadoopRDD.ignoreEmptySplits` sur `false`. La EMR version 6.9.0 d'Amazon résout ce problème.
- Sur les clusters de longue durée de Trino, Amazon EMR 6.6.0 active les paramètres de journalisation de Garbage Collection dans le fichier `jvm.config` de Trino afin d'obtenir de meilleures informations à partir des journaux de Garbage Collection. Cette modification ajoute de nombreux journaux du récupérateur de mémoire au fichier `launcher.log` (`/var/log/trino/launcher.log`). Si vous utilisez des clusters Trino dans Amazon EMR 6.6.0, vous risquez de rencontrer des nœuds à court d'espace disque après quelques jours d'exécution du cluster en raison des journaux ajoutés.

La solution à ce problème consiste à exécuter le script ci-dessous en tant qu'action Bootstrap afin de désactiver les paramètres de journalisation de Garbage Collection dans `jvm.config` lors de la création ou du clonage du cluster pour Amazon 6.6.0. EMR

```
#!/bin/bash
```

```
set -ex
PRESTO_PUPPET_DIR='/var/aws/emr/bigtop-deploy/puppet/modules/trino'
sudo bash -c "sed -i '/-Xlog/d' ${PRESTO_PUPPET_DIR}/templates/jvm.config"
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration

`spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

- Avec les EMR versions 5.36.0 et 6.6.0 à 6.9.0 d'Amazon, `SecretAgent` les composants du `RecordServer` service peuvent subir une perte de données de journal en raison d'une configuration incorrecte du modèle de nom de fichier dans les propriétés `Log4j2`. En cas de configuration incorrecte, les composants ne génèrent qu'un seul fichier journal par jour. Lorsque la stratégie de rotation est appliquée, elle remplace le fichier existant au lieu de générer un nouveau fichier journal comme prévu. Pour contourner le problème, utilisez une action d'amorçage pour générer des journaux toutes les heures et ajoutez un nombre entier auto-incrémenté dans le nom du fichier pour gérer la rotation.

Pour les versions Amazon EMR 6.6.0 à 6.9.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Pour Amazon EMR 5.36.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Versions des composants 6.6.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	3.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-notebook-env	1.5.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.20.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.1.0	EMRConnecteur S3Select
emrfs	2,50,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.14.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.14.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	3.2.1-amzn-6	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-6	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-6	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-6	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-6	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.2.1-amzn-6	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-6	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-6	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-6	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	3.2.1-amzn-6	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-6	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.4-amzn-2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.4-amzn-2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.4.4-amzn-2	HBaseclient en ligne de commande.
hbase-rest-server	2.4.4-amzn-2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.4-amzn-2	Service fournissant un point de terminaison Thrift pour. HBase
hbase-operator-tools	2.4.4-amzn-2	Outil de réparation pour les HBase clusters Apache.
hcatalog-client	3.1.2-amzn-7	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	3.1.2-amzn-7	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-7	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.2-amzn-7	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-7	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-7	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-7	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.10.1-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.10.1-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.10.1-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.

Composant	Version	Description
hudi-spark	0.10.1-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	0.13.1	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.8.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	11,0,194	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-connectors	5.1.2	Connecteurs Apache Phoenix pour Spark-3
phoenix-query-server	5.1.2	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.267-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.267-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.267-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	367-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.

Composant	Version	Description
trino-worker	367-amzn-0	Service pour exécuter les éléments d'une requête.
trino-client	367-amzn-0	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.2.0-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.2.0-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.2.0-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.2.0-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	22.02.0-amzn-0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.

Composant	Version	Description
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0.10.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.7	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.7	ZooKeeper client en ligne de commande.

Classifications des configurations 6.6.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration

EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.6.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode,

Classifications	Description	Actions de reconfiguration
		SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.

Classifications	Description	Actions de reconfiguration
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.

Classifications	Description	Actions de reconfiguration
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-tpcds	Modifiez les valeurs dans le fichier <code>tpcds.properties</code> de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier <code>trino-env.sh</code> de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Modifiez les valeurs dans le fichier <code>iceberg.properties</code> de Trino.	Restarts Trino-Server (for Trino)

Classifications	Description	Actions de reconfiguration
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.

Classifications	Description	Actions de reconfiguration
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.

Classifications	Description	Actions de reconfiguration
yarn-env	Modifiez les valeurs dans l'YARN environnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 6.5.0

Versions d'application 6.5.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
Kit AWS SDK pour Java	1,12,31	1,12,31	1,1,977	1,1,977
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.0	1.13.1	1.12.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.4.4	2.2.6	2.2.6
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.9.0-amzn-1	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
Hue	4.9.0	4.9.0	4.9.0	4.9.0
Iceberg	0.12.0	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	5.1.2	5.1.2	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,261	0,254,1	0,245,1	0,245,1
Spark	3.1.2	3.1.2	3.1.1	3.1.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	360	359	350	350
Zeppelin	0.10.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.5.7	3.5.7	3.4.14	3.4.14

Notes de mise à jour 6.5.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.5.0 d'Amazon. Les modifications ont été apportées à la version 6.4.0.

Date de parution initiale : 20 janvier 2022

Date de mise à niveau : 21 mars 2022

Nouvelles fonctions

- [Dimensionnement géré] Optimisation du dimensionnement géré par les données Spark Shuffle - Pour EMR les versions 5.34.0 et ultérieures d'Amazon, ainsi que pour EMR les versions 6.4.0 et ultérieures, le dimensionnement géré prend désormais en compte les données Spark shuffle (données que Spark redistribue entre les partitions pour effectuer des opérations spécifiques). Pour plus d'informations sur les opérations de shuffle, consultez la section [Utilisation du dimensionnement EMR géré EMR dans Amazon](#) dans le Guide de EMRgestion Amazon et le Guide de [programmation Spark](#).
- À partir d'Amazon EMR 5.32.0 et 6.5.0, le dimensionnement dynamique de l'exécuteur pour Apache Spark est activé par défaut. Pour activer ou désactiver cette fonctionnalité, vous pouvez utiliser le paramètre de configuration `spark.yarn.heterogeneousExecutors.enabled`.
- Prise en charge du format de table ouvert Apache Iceberg pour les jeux de données analytiques volumineux.
- Support pour ranger-trino-plugin 2.0.1-amzn-1
- Prise en charge de toree 0.5.0

Modifications, améliorations et problèmes résolus

- La version EMR 6.5 d'Amazon prend désormais en charge Apache Iceberg 0.12.0 et fournit des améliorations d'exécution avec Amazon EMR Runtime pour Apache Spark, Amazon EMR Runtime pour Presto et Amazon EMR Runtime pour Apache Hive.
- [Apache Iceberg](#) est un format de table ouvert pour les grands ensembles de données dans Amazon S3. Il fournit des performances de requête rapides sur de grandes tables, des validations atomiques, des écritures simultanées et une évolution de table SQL compatible. Avec la EMR version 6.5, vous pouvez utiliser Apache Spark 3.1.2 avec le format de table Iceberg.

- Apache Hudi 0.9 ajoute Spark SQL DDL et le DML support. Cela vous permet de créer et de modifier des tables Hudi en utilisant uniquement SQL des instructions. Apache Hudi 0.9 inclut également des améliorations des performances côté requête et côté écriture.
- Amazon EMR Runtime pour Apache Hive améliore les performances d'Apache Hive sur Amazon S3 en supprimant les opérations de renommage lors des opérations intermédiaires, et améliore les performances des commandes metastore check (MSCK) utilisées pour réparer les tables.

Problèmes connus

- Lorsque les EMR versions 6.5.0, 6.6.0 ou 6.7.0 d'Amazon lisent les tables Apache Phoenix via le shell Apache Spark, cela se produit `NoSuchMethodError` car Amazon EMR utilise une valeur incorrecte. `Hbase.compat.version` La EMR version 6.8.0 d'Amazon résout ce problème.
- Les clusters de la solution groupée Hbase en haute disponibilité (HA) ne parviennent pas à se provisionner avec la taille de volume et le type d'instance par défaut. La solution à ce problème consiste à augmenter la taille du volume racine.
- Pour utiliser les actions Spark avec Apache Oozie, vous devez ajouter la configuration suivante à votre fichier Oozie `workflow.xml`. Sinon, plusieurs bibliothèques critiques telles que Hadoop et B EMRFS seront absentes du classpath des exécuteurs Spark lancés par Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-`. Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 6.5.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	3.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-notebook-env	1.4.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.19.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.1.0	EMRConnecteur S3Select
emrfs	2,48,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.14.0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.14.0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	3.2.1-amzn-5	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-5	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-5	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-5	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-5	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.2.1-amzn-5	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-5	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-5	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-5	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	3.2.1-amzn-5	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-5	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.4-amzn-1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.4-amzn-1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.4.4-amzn-1	HBaseclient en ligne de commande.
hbase-rest-server	2.4.4-amzn-1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.4.4-amzn-1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	3.1.2-amzn-6	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	3.1.2-amzn-6	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-6	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.2-amzn-6	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-6	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-6	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-6	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.9.0-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.9.0-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.9.0-amzn-1	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.

Composant	Version	Description
hudi-spark	0.9.0-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.9.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
iceberg	0.12.0	Apache Iceberg est un format de table ouvert pour les jeux de données analytiques de grande taille.
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.8.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	5.1.2	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.261-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.261-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.261-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	360	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	360	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
trino-client	360	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.1.2-amzn-1	Clients de ligne de commande Spark.
spark-history-server	3.1.2-amzn-1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.1.2-amzn-1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.1.2-amzn-1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	0,4.1	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0.10.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.7	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.7	ZooKeeper client en ligne de commande.

Classifications des configurations 6.5.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.5.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classifications	Description	Actions de reconfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
iceberg-defaults	Modification des valeurs du fichier iceberg-defaults.conf d'Iceberg.	Not available.

Classifications	Description	Actions de reconfiguration
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.

Classifications	Description	Actions de reconfiguration
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-tpcds	Modifiez les valeurs dans le fichier <code>tpcds.properties</code> de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier <code>trino-env.sh</code> de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Trino.	Not available.
trino-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 6.4.0

Versions d'application 6.4.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)

- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
Kit AWS SDK pour Java	1,12,31	1,1,977	1,1,977	1,1880
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.13.1	1.12.1	1.12.1	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.4.4	2.2.6	2.2.6	2.2.6-amzn-0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.8.0-amzn-0	0.7.0-amzn-0	0.7.0-amzn-0	0.6.0-amzn-1
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.1	5.2.0
Phoenix	5.1.2	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,254,1	0,245,1	0,245,1	0,238,3
Spark	3.1.2	3.1.1	3.1.1	3.0.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	359	350	350	343
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.5.7	3.4.14	3.4.14	3.4.14

Notes de mise à jour 6.4.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.4.0 d'Amazon. Les modifications ont été apportées à la version 6.3.0.

Date de parution initiale : 20 septembre 2021

Date de mise à niveau : 21 mars 2022

Applications prises en charge

- AWS SDK for Java version 1.12.31
- CloudWatch Version 2.2.0 de l'évier
- Connecteur DynamoDB version 4.16.0
- EMRFSversion 2.47.0
- Amazon EMR Goodies version 3.2.0
- Connecteur Amazon EMR Kinesis version 3.5.0
- Amazon EMR Record Server version 2.1.0
- Amazon EMR Scripts version 2.5.0
- Flink version 1.13.1
- Ganglia version 3.7.2
- AWS Client Glue Hive Metastore version 3.3.0
- Hadoop version 3.2.1-amzn-4
- HBaseversion 2.4.4-amzn-0
- HBase-operator-tools 1.1.0
- HCatalogversion 3.1.2-amzn-5
- Hive version 3.1.2-amzn-5
- Hudi version 0.8.0-amzn-0
- Hue version 4.9.0
- JDKVersion Java Corretto-8.302.08.1 (version 1.8.0_302-b08)
- JupyterHub version 1.4.1
- Livy version 0.7.1-incubating
- MXNetversion 1.8.0
- Oozie version 5.2.1
- Phoenix version 5.1.2
- Pig version 0.17.0
- Presto version 0.254.1-amzn-0
- Trino version 359
- Apache Ranger KMS (chiffrement transparent multi-maîtres) version 2.0.0

- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.2.0
- SageMaker SDKVersion 1.4.1 de Spark
- Scala version 2.12.10 (machine virtuelle de serveur ouverte JDK 64 bits, Java 1.8.0_282)
- Spark version 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop version 1.4.7
- TensorFlow version 2.4.1
- tez version 0.9.2
- Zeppelin version 0.9.0
- Zookeeper version 3.5.7
- Connecteurs et pilotes : Connecteur DynamoDB 4.16.0

Nouvelles fonctionnalités

- [Dimensionnement géré] Optimisation du dimensionnement géré par les données Spark Shuffle - Pour EMR les versions 5.34.0 et ultérieures d'Amazon, ainsi que pour EMR les versions 6.4.0 et ultérieures, le dimensionnement géré prend désormais en compte les données Spark shuffle (données que Spark redistribue entre les partitions pour effectuer des opérations spécifiques). Pour plus d'informations sur les opérations de shuffle, consultez la section [Utilisation du dimensionnement EMR géré EMR dans Amazon](#) dans le Guide de EMRgestion Amazon et le Guide de [programmation Spark](#).
- Sur les EMR clusters Amazon compatibles avec Apache Ranger, vous pouvez utiliser Apache Spark SQL pour insérer des données dans les tables de métastore Apache Hive ou les mettre à jour à l'aide de, et. INSERT INTO INSERT OVERWRITE ALTER TABLE Lors de l'utilisation ALTER TABLE avec SparkSQL, l'emplacement d'une partition doit être le répertoire enfant d'un emplacement de table. Amazon EMR ne prend actuellement pas en charge l'insertion de données dans une partition où l'emplacement de la partition est différent de celui de la table.
- Presto SQL a été [renommé Trino](#).
- Hive : L'exécution de SELECT requêtes simples avec LIMIT clause est accélérée en arrêtant l'exécution de la requête dès que le nombre d'enregistrements mentionnés dans la LIMIT clause est récupéré. Les SELECT requêtes simples sont des requêtes qui ne contiennent pas de clause GROUP BY/ORDERby ou des requêtes qui n'ont pas d'étape réductrice. Par exemple, SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>.

Contrôles de simultanéité Hudi

- Hudi prend désormais en charge Optimistic Concurrency Control (OCC), qui peut être utilisé avec des opérations d'écriture telles que UPSERT et INSERT pour autoriser les modifications de plusieurs rédacteurs vers la même table Hudi. C'est au niveau du fichierOCC, donc deux validations (ou auteurs) peuvent écrire dans la même table, si leurs modifications ne sont pas en conflit. Pour plus d'informations, consultez le [contrôle de simultanéité de Hudi](#).
- Zookeeper est installé sur les EMR clusters Amazon, qui peut être utilisé comme fournisseur de verrous pour OCC. Pour faciliter l'utilisation de cette fonctionnalité, les propriétés préconfigurées des EMR clusters Amazon sont les suivantes :

```
hoodie.write.lock.provider=org.apache.hudi.client.transaction.lock.ZookeeperBasedLockProvider
hoodie.write.lock.zookeeper.url=<EMR Zookeeper URL>
hoodie.write.lock.zookeeper.port=<EMR Zookeeper Port>
hoodie.write.lock.zookeeper.base_path=/hudi
```

Pour les activerOCC, vous devez configurer les propriétés suivantes soit avec leurs options de tâche Hudi, soit au niveau du cluster à l'aide des configurations Amazon : EMR API

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily
instead of inline with every write)
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table
Name is a good option)
```

Hudi Monitoring : CloudWatch intégration d'Amazon pour générer des rapports sur Hudi Metrics

- Amazon EMR prend en charge la publication de Hudi Metrics sur Amazon CloudWatch. Elle est activée en définissant les configurations requises suivantes :

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- Les configurations Hudi facultatives que vous pouvez modifier sont les suivantes :

Paramètre	Description	Valeur
<code>hoodie.metrics.cloudwatch.reporting.period.seconds</code>	Fréquence (en secondes) à laquelle les statistiques doivent être communiquées à Amazon CloudWatch	La valeur par défaut est de 60 s, ce qui convient à la résolution d'une minute par défaut proposée par Amazon CloudWatch
<code>hoodie.metrics.cloudwatch.metric.prefix</code>	Préfixe à ajouter au nom de chaque métrique	La valeur par défaut est vide (pas de préfixe)
<code>hoodie.metrics.cloudwatch.namespaces</code>	CloudWatch Espace de noms Amazon sous lequel les métriques sont publiées	La valeur par défaut est Hudi
<code>hoodie.metrics.cloudwatch.maxDatumsPerDemande</code>	Nombre maximum de références à inclure dans une demande adressée à Amazon CloudWatch	La valeur par défaut est 20, identique à la valeur CloudWatch par défaut d'Amazon

Support et améliorations des configurations Amazon EMR Hudi

- Les clients peuvent désormais tirer parti EMR de la fonctionnalité de configuration API et de reconfiguration pour configurer les configurations Hudi au niveau du cluster. Une nouvelle prise en charge de la configuration basée sur les fichiers a été introduite via `/etc/hudi/conf/hudi-defaults.conf`, à l'instar d'autres applications telles que Spark, Hive, etc. EMRconfigure quelques paramètres par défaut pour améliorer l'expérience utilisateur :

— `hoodie.datasource.hive_sync.jdbcurl` est configuré sur le serveur Hive du cluster URL et n'a plus besoin d'être spécifié. Cela est particulièrement utile lorsque vous exécutez une tâche en mode cluster Spark, où vous deviez auparavant spécifier l'adresse IP EMR principale d'Amazon.

— HBase des configurations spécifiques, utiles pour utiliser l'HBaseindex avec Hudi.

— Configuration spécifique au fournisseur de verrous Zookeeper, comme indiqué dans la section Contrôle de simultanéité, qui facilite l'utilisation d'Optimistic Concurrency Control (). OCC

- Des modifications supplémentaires ont été introduites pour réduire le nombre de configurations à transmettre et pour en déduire automatiquement dans la mesure du possible :
 - Le mot clé `partitionBy` peut être utilisé pour spécifier la colonne de partition.
 - Lorsque vous activez Hive Sync, il n'est plus obligatoire de spécifier `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY`. Ces valeurs peuvent être déduites du nom de la table Hudi et du champ de partition.
 - `KEYGENERATOR_CLASS_OPT_KEY` n'est pas obligatoire et peut être déduite de cas plus simples de `SimpleKeyGenerator` et `ComplexKeyGenerator`.

Mises en garde de Hudi

- Hudi ne prend pas en charge l'exécution vectorisée dans les tables Hive for Merge on Read (MoR) et Bootstrap. Par exemple, `count(*)` échoue avec la table en temps réel de Hudi lorsque `hive.vectorized.execution.enabled` est défini sur `true`. Comme solution de contournement, vous pouvez désactiver la lecture vectorisée en définissant `hive.vectorized.execution.enabled` sur `false`.
- La prise en charge des enregistreurs multiples n'est pas compatible avec la fonction d'amorçage de Hudi.
- Flink Streamer et Flink SQL sont des fonctionnalités expérimentales de cette version. Ces fonctionnalités ne sont pas recommandées pour les déploiements de production.

Modifications, améliorations et problèmes résolus

Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.

- Auparavant, le redémarrage manuel du gestionnaire de ressources sur un cluster multi-maîtres obligeait les EMR démons Amazon du cluster, tels que Zookeeper, à recharger tous les nœuds précédemment désaffectés ou perdus dans le fichier `znode` de Zookeeper. Cela a entraîné le dépassement des limites par défaut dans certaines situations. Amazon supprime EMR désormais les enregistrements de nœuds mis hors service ou perdus datant de plus d'une heure du fichier Zookeeper et les limites internes ont été augmentées.

- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une situation de course lors de la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Configuration d'un cluster pour résoudre les problèmes de performances du serveur Apache YARN Timeline versions 1 et 1.5

Les versions 1 et 1.5 d'Apache YARN Timeline Server peuvent entraîner des problèmes de performances avec de grands EMR clusters très actifs `yarn.resourcemanager.system-metrics-publisher.enabled=true`, en particulier avec le paramètre par défaut d'AmazonEMR. Un YARN Timeline Server v2 open source résout le problème de performance lié à l'évolutivité du YARN Timeline Server.

Les autres solutions à ce problème incluent :

- Configuration de `yarn.resourcemanager.system-metrics-publisher.enabled=false` dans le fichier `yarn-site.xml`.
- Activation du correctif pour ce problème lors de la création d'un cluster, comme décrit ci-dessous.

Les EMR versions Amazon suivantes contiennent un correctif pour ce problème de performance du serveur YARN Timeline.

EMR5,30.2, 5,31.1, 5,32.1, 5,33.1, 5,34.x, 6,0.1, 6.1.1, 6.2.1, 6.3.1, 6.4.x

Pour activer le correctif sur l'une des EMR versions Amazon spécifiées ci-dessus, définissez ces propriétés sur `true` dans un JSON fichier de configuration transmis à l'aide du [paramètre de aws emr create-cluster commande](#) :`--configurations file://./configurations.json`. Vous pouvez également activer le correctif à l'aide de l'[interface utilisateur de la console de reconfiguration](#).

Exemple du contenu du fichier `configurations.json` :

```
[
{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
      "true",
    "yarn.resourcemanager.system-metrics-publisher.enabled": "true"
  },
  "Configurations": []
}
```

- Le Web HDFS et le serveur HTTPFS sont désactivés par défaut. Vous pouvez réactiver le Web à HDFS l'aide de la configuration Hadoop, `dfs.webhdfs.enabled`. Le serveur HttpFS peut être démarré en utilisant `sudo systemctl start hadoop-httpfs`.
- HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPCE politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages](#).
- Hive : les performances des requêtes d'écriture sont améliorées en permettant l'utilisation d'un répertoire temporaire activé HDFS pour la dernière tâche. Les données temporaires pour la tâche finale sont écrites sur Amazon S3 HDFS plutôt que sur Amazon S3 et les performances sont

améliorées car les données sont déplacées depuis HDFS l'emplacement de la table finale (Amazon S3) plutôt qu'entre les appareils Amazon S3.

- Hive : amélioration du temps de compilation des requêtes jusqu'à 2,5 fois avec l'élimination des partitions du métastore Glue.
- Par défaut, lorsque les éléments intégrés UDFs sont transmis par Hive au serveur Hive Metastore, seul un sous-ensemble de ces éléments intégrés UDFs est transmis au Glue Metastore, car Glue ne prend en charge que des opérateurs d'expression limités. Si vous définissez `hive.glue.partition.pruning.client=true`, tout l'élimination des partitions se fait du côté client. Si vous définissez `hive.glue.partition.pruning.server=true`, tout l'élimination des partitions se fait du côté serveur.

Problèmes connus

- Les requêtes Hue ne fonctionnent pas dans Amazon EMR 6.4.0 car le serveur Apache Hadoop HTTPFS est désactivé par défaut. Pour utiliser Hue on Amazon EMR 6.4.0, démarrez manuellement le serveur HTTPFS sur le nœud EMR principal Amazon à l'aide `sudo systemctl start hadoop-httpfs` ou [utilisez une étape Amazon EMR](#).
- La fonctionnalité Amazon EMR Notebooks utilisée avec l'usurpation d'identité d'un utilisateur de Livy ne fonctionne pas car le protocole HTTPFS est désactivé par défaut. Dans ce cas, le EMR bloc-notes ne peut pas se connecter au cluster sur lequel l'emprunt d'identité de Livy est activé. La solution consiste à démarrer le serveur HTTPFS avant de connecter le EMR bloc-notes au cluster à l'aide de `sudo systemctl start hadoop-httpfs`
- Dans la EMR version 6.4.0 d'Amazon, Phoenix ne prend pas en charge le composant des connecteurs Phoenix.
- Pour utiliser les actions Spark avec Apache Oozie, vous devez ajouter la configuration suivante à votre fichier Oozie `workflow.xml`. Sinon, plusieurs bibliothèques critiques telles que Hadoop et BMRFS seront absentes du classpath des exécuteurs Spark lancés par Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.

- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 6.4.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-goodies	3.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.3.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.18.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.1.0	EMRConnecteur S3Select
emrfs	2,47,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.13.1	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.13.1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.2.1-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	3.2.1-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.2.1-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.4.4-amzn-0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.4.4-amzn-0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.4.4-amzn-0	HBaseclient en ligne de commande.
hbase-rest-server	2.4.4-amzn-0	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	2.4.4-amzn-0	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	3.1.2-amzn-5	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.2-amzn-5	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-5	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	3.1.2-amzn-5	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-5	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-5	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-5	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.8.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.

Composant	Version	Description
hudi-presto	0.8.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-trino	0.8.0-amzn-0	Bibliothèque de solution groupée pour exécuter Trino avec Hudi.
hudi-spark	0.8.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.9.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.8.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5.1.2	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	5.1.2	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.254.1-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.254.1-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.254.1-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
trino-coordinator	359	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants trino-workers.
trino-worker	359	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
trino-client	359	Client de ligne de commande Trino installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Trino n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.1.2-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.1.2-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.1.2-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.1.2-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	0,4.1	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0.9.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.5.7	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.5.7	ZooKeeper client en ligne de commande.

Classifications des configurations 6.4.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.4.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classifications	Description	Actions de reconfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.

Classifications	Description	Actions de reconfiguration
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l'application MapReduce.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier <code>oozie-log4j.properties</code> d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier <code>oozie-site.xml</code> d'Oozie.	Restarts Oozie.

Classifications	Description	Actions de reconfiguration
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
trino-log	Modifiez les valeurs dans le fichier log.properties de Trino.	Restarts Trino-Server (for Trino)

Classifications	Description	Actions de reconfiguration
trino-config	Modifiez les valeurs dans le fichier config.properties de Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Trino.	Restarts Trino-Server (for Trino)
trino-env	Modification des valeurs dans le fichier trino-env.sh de Trino.	Restarts Trino-Server (for Trino)
trino-node	Modifiez les valeurs dans le fichier node.properties de Trino.	Not available.
trino-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Trino.	Not available.
trino-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Trino.	Not available.
trino-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Trino.	Not available.
trino-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Trino.	Not available.
trino-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
trino-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Trino.	Not available.
trino-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Trino.	Not available.
trino-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Trino.	Not available.
trino-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Trino.	Not available.
trino-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Trino.	Not available.
trino-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Trino.	Not available.
trino-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Trino.	Not available.
trino-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Trino.	Not available.
trino-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Trino.	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 6.3.1

Versions d'application 6.3.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
Kit AWS SDK pour Java	1,1,977	1,1,977	1,1880	1,1880
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.7.0-amzn-0	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.2.2	1.2.2	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.7.0

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,245,1	0,245,1	0,238,3	0,238,3
Spark	3.1.1	3.1.1	3.0.1	3.0.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	350	350	343	343
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 6.3.1

Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.

Modifications, améliorations et problèmes résolus

- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR

- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une situation de course lors de la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPCE politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages.](#)

Problèmes connus

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.

- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 6.3.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	3.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.2.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.18.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.1.0	EMRConnecteur S3Select
emrfs	2,46,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.12.1	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.12.1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-3.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-3.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-3.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-3.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-3.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.2.1-amzn-3.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-3.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-3.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	3.2.1-amzn-3.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.2.1-amzn-3.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-3.1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.2.6-amzn-1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.2.6-amzn-1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.2.6-amzn-1	HBaseclient en ligne de commande.
hbase-rest-server	2.2.6-amzn-1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.2.6-amzn-1	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	3.1.2-amzn-4	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	3.1.2-amzn-4	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-4	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.2-amzn-4	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-4	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-4	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-4	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.7.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.7.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-prestosql	0.7.0-amzn-0	Bibliothèque de bundles pour exécuter Presto SQL avec Hudi.

Composant	Version	Description
hudi-spark	0.7.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.9.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.2.2	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.7.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur

Composant	Version	Description
phoenix-library	5,0,0- -2,0 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	5,0,0- -2,0 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.245.1-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.245.1-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.245.1-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
prestosql-coordinator	350	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants prestosql-workers.
prestosql-worker	350	Service pour exécuter les éléments d'une requête.
prestosql-client	350	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.1.1-amzn-0.1	Clients de ligne de commande Spark.
spark-history-server	3.1.1-amzn-0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.1.1-amzn-0.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.1.1-amzn-0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	0,4.1	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.

Composant	Version	Description
tez-on-yarn	0.9.2	L'YARN Application Tez et les bibliothèques.
webserver	2.4.41+	HTTPserveur Apache.
zeppelin-server	0.9.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 6.3.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.3.1

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the Resource Manager service.

Classifications	Description	Actions de reconfiguration
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.

Classifications	Description	Actions de reconfiguration
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.

Classifications	Description	Actions de reconfiguration
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
prestosql-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoSQL)

Classifications	Description	Actions de reconfiguration
prestosql-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Modifiez les valeurs dans le fichier SQL node.properties de Presto.	Not available.
prestosql-connector-blackhole	Modifiez les valeurs dans le fichier SQL blackhole.properties de Presto.	Not available.
prestosql-connector-cassandra	Modifiez les valeurs dans le fichier SQL cassandra.properties de Presto.	Not available.
prestosql-connector-hive	Modifiez les valeurs dans le fichier SQL hive.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Modifiez les valeurs dans le fichier SQL jmx.properties de Presto.	Not available.
prestosql-connector-kafka	Modifiez les valeurs dans le fichier SQL kafka.properties de Presto.	Not available.
prestosql-connector-localfile	Modifiez les valeurs dans le fichier SQL localfile.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
prestosql-connector-memory	Modifiez les valeurs dans le fichier SQL memory.properties de Presto.	Not available.
prestosql-connector-mongodb	Modifiez les valeurs dans le fichier SQL mongodb.properties de Presto.	Not available.
prestosql-connector-mysql	Modifiez les valeurs dans le fichier SQL mysql.properties de Presto.	Not available.
prestosql-connector-postgresql	Modifiez les valeurs dans le fichier SQL postgresql.properties de Presto.	Not available.
prestosql-connector-raptor	Modifiez les valeurs dans le fichier SQL raptor.properties de Presto.	Not available.
prestosql-connector-redis	Modifiez les valeurs dans le fichier SQL redis.properties de Presto.	Not available.
prestosql-connector-redshift	Modifiez les valeurs dans le fichier SQL redshift.properties de Presto.	Not available.
prestosql-connector-tpch	Modifiez les valeurs dans le fichier SQL tpch.properties de Presto.	Not available.
prestosql-connector-tpcds	Modifiez les valeurs dans le fichier SQL tpcds.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARN environnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 6.3.0

Versions d'application 6.3.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
Kit AWS SDK pour Java	1,1,977	1,1880	1,1880	1,1,828
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.7.0-amzn-0	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-incubating-amzn-2
Hue	4.9.0	4.8.0	4.8.0	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	-
JupyterHub	1.2.2	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
MXNet	1.7.0	1.7.0	1.7.0	1.6.0
Mahout	-	-	-	-
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,245,1	0,238,3	0,238,3	0.232
Spark	3.1.1	3.0.1	3.0.1	3.0.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.3.1	2.3.1	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	350	343	343	338
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 6.3.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.3.0 d'Amazon. Les modifications ont été apportées à la version 6.2.0.

Date de parution initiale : 12 mai 2021

Dernière mise à jour : 9 août 2021

Applications prises en charge

- AWS SDK for Java version 1.11.977

- CloudWatch Version 2.1.0 de l'évier
- Connecteur DynamoDB version 4.16.0
- EMRFSversion 2.46.0
- Amazon EMR Goodies version 3.2.0
- Connecteur Amazon EMR Kinesis version 3.5.0
- Amazon EMR Record Server version 2.0.0
- Amazon EMR Scripts version 2.5.0
- Flink version 1.12.1
- Ganglia version 3.7.2
- AWS Client Glue Hive Metastore version 3.2.0
- Hadoop version 3.2.1-amzn-3
- HBaseversion 2.2.6-amzn-1
- HBase-operator-tools 1.0.0
- HCatalogversion 3.1.2-amzn-0
- Hive version 3.1.2-amzn-4
- Hudi version 0.7.0-amzn-0
- Hue version 4.9.0
- JDKVersion Java Corretto-8.282.08.1 (version 1.8.0_282-b08)
- JupyterHub version 1.2.0
- Livy version 0.7.0-incubating
- MXNetversion 1.7.0
- Oozie version 5.2.1
- Phoenix version 5.0.0
- Pig version 0.17.0
- Presto version 0.245.1-amzn-0
- Presto SQL version 350
- Apache Ranger KMS (chiffrement transparent multi-maîtres) version 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.1.0

- SageMaker SDKVersion 1.4.1 de Spark
- Scala version 2.12.10 (machine virtuelle de serveur ouverte JDK 64 bits, Java 1.8.0_282)
- Spark version 3.1.1-amzn-0
- spark-rapids 0.4.1
- Sqoop version 1.4.7
- TensorFlow version 2.4.1
- tez version 0.9.2
- Zeppelin version 0.9.0
- Zookeeper version 3.4.14
- Connecteurs et pilotes : Connecteur DynamoDB 4.16.0

Nouvelles fonctionnalités

- Amazon EMR prend en charge les points d'accès Amazon S3, une fonctionnalité d'Amazon S3 qui vous permet de gérer facilement l'accès aux lacs de données partagés. En utilisant votre alias Amazon S3 Access Point, vous pouvez simplifier l'accès aux données à grande échelle sur AmazonEMR. Vous pouvez utiliser les points d'accès Amazon S3 avec toutes les versions d'Amazon EMR sans frais supplémentaires dans toutes les AWS régions où Amazon EMR est disponible. Pour en savoir plus sur les points d'accès Amazon S3 et les alias de points d'accès, consultez [Utilisation d'un alias de type compartiment pour votre point d'accès](#) dans le Guide de l'utilisateur Amazon S3.
- Les nouveaux `DescribeReleaseLabel` `ListReleaseLabel` API paramètres fournissent des informations sur les étiquettes EMR de sortie d'Amazon. Vous pouvez répertorier par programmation les versions disponibles dans la région où la API demande est exécutée et répertorier les applications disponibles pour une étiquette de EMR version Amazon spécifique. Les paramètres des étiquettes de version répertorient également les EMR versions d'Amazon qui prennent en charge une application spécifique, telle que Spark. Ces informations peuvent être utilisées pour lancer des clusters Amazon EMR par programmation. Par exemple, vous pouvez lancer un cluster à l'aide de la dernière version des résultats de `ListReleaseLabel`. Pour plus d'informations, consultez [DescribeReleaseLabel](#) et consultez [ListReleaseLabels](#) le Amazon EMR API Reference.
- Avec Amazon EMR 6.3.0, vous pouvez lancer un cluster qui s'intègre nativement à Apache Ranger. Apache Ranger est un cadre open source permettant d'activer, de surveiller et de gérer la sécurité globale des données sur la plateforme Hadoop. Pour plus d'informations, consultez

[Apache Ranger](#). Grâce à l'intégration native, vous pouvez utiliser votre propre Apache Ranger pour appliquer un contrôle précis de l'accès aux données sur Amazon. EMR Consultez la section [Intégrer Amazon EMR à Apache Ranger](#) dans le guide EMR de gestion Amazon.

- Politiques gérées délimitées : pour s'aligner sur les AWS meilleures pratiques, Amazon EMR a introduit des politiques gérées par défaut de EMR portée v2 en remplacement des politiques qui seront obsolètes. Consultez les [politiques EMR gérées par Amazon](#).
- État du support du service de métadonnées d'instance (IMDS) V2 : pour Amazon EMR 6.2 ou version ultérieure, les EMR composants Amazon sont utilisés IMDSv2 pour tous les IMDS appels. Pour les IMDS appels dans le code de votre application, vous pouvez utiliser les deux IMDSv1 et IMDSv2, ou les configurer pour qu'IMDS ils ne soient utilisés que IMDSv2 pour renforcer la sécurité. Si vous la désactivez IMDSv1 dans les versions antérieures d'Amazon EMR 6.x, cela entraîne un échec du démarrage du cluster.

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une situation de course lors de la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification

Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.

- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- La valeur par défaut du mode d'explication de l'SQLinterface utilisateur de Spark extended est passée de à formatted dans [Spark 3.1](#). Amazon l'EMR a rétabli extended pour inclure les informations du plan logique dans l'SQLinterface utilisateur de Spark. Ceci peut être inversé en définissant `spark.sql.ui.explainMode` sur formatted.
- Les commits suivants ont été rétroportés depuis la branche master de Spark.
 - [\[SPARK-34752\]](#) [\[BUILD\]](#) [Passez](#) Jetty au 9.4.37 à l'adresse -2020-27223. CVE
 - [\[SPARK-34534\]](#) Corrige l' blockIds ordre lors de l'utilisation FetchShuffleBlocks pour récupérer des blocs.
 - [\[SPARK-34681\]](#) [\[SQL\]](#) Correction d'un bogue concernant la jointure par hachage externe complète lors de la construction du côté gauche avec des conditions inégales.
 - [\[SPARK-34497\]](#) [\[SQL\]](#) Corrigez les fournisseurs de JDBC connexion intégrés pour rétablir les modifications du contexte de JVM sécurité.
- Pour améliorer l'interopérabilité avec le RAPIDS plugin Nvidia Spark, une solution de contournement a été ajoutée pour résoudre un problème empêchant le déclenchement de l'élagage dynamique des partitions lors de l'utilisation de Nvidia Spark RAPIDS avec l'exécution adaptative des requêtes désactivée, voir les [RAPIDSproblèmes #1378](#) et [# RAPIDS #1386](#). Pour plus de détails sur la nouvelle configurationspark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse, consultez le [RAPIDSnuméro #1386](#).
- L'algorithme par défaut du validateur de sortie de fichier a été modifié de l'algorithme v2 à l'algorithme v1 dans Spark 3.1 open source. Pour plus d'informations, consultez cet article [Amazon EMR optimisant les performances de Spark : élagage dynamique des partitions](#).
- Amazon EMR est revenu à l'algorithme v2, l'algorithme par défaut utilisé dans les versions précédentes d'Amazon EMR 6.x, pour empêcher la régression des performances. Pour rétablir le comportement de Spark 3.1 open source, définissez

`spark.hadoop.mapreduce.fileoutputcommitter.algorithm.version` sur 1. Le Spark open source a apporté cette modification parce que la validation des tâches dans l'algorithme v2 du validateur de sortie de fichier n'est pas atomique, ce qui peut entraîner un problème d'exactitude des données de sortie dans certains cas. Cependant, la validation des tâches dans l'algorithme v1 n'est pas non plus atomique. Dans certains scénarios, la validation des tâches comprend une suppression effectuée avant un changement de nom. Cela peut entraîner un problème silencieux d'exactitude des données.

- Les problèmes de Managed Scaling rencontrés dans les EMR versions précédentes d'Amazon ont été corrigés et des améliorations ont été apportées afin de réduire de manière significative les taux d'échec des applications.
- Le AWS Java SDK Bundle a été installé sur chaque nouveau cluster. Il s'agit d'un fichier jar unique contenant tous les services SDKs et leurs dépendances, au lieu de fichiers jar de composants individuels. Pour plus d'informations, consultez [Java SDK Bundled Dependency](#).

Problèmes connus

- Pour les clusters de sous-réseaux privés Amazon EMR 6.3.0 et 6.2.0, vous ne pouvez pas accéder à l'interface utilisateur Web de Ganglia. Vous recevrez un message d'erreur « accès refusé (403) ». D'autres sites WebUIs, tels que Spark, Hue JupyterHub, Zeppelin, Livy et Tez, fonctionnent normalement. L'accès à l'interface utilisateur Web de Ganglia sur les clusters de sous-réseaux publics fonctionne également normalement. Pour résoudre ce problème, redémarrez le service `httpd` sur le nœud primaire avec `sudo systemctl restart httpd`. Ce problème est résolu dans Amazon EMR 6.4.0.
- Lorsque AWS Glue Data Catalog est activé, l'utilisation de Spark pour accéder à une base de données AWS Glue avec un emplacement de chaîne nul URI peut échouer. Cela se produit pour les EMR versions antérieures d'Amazon, mais avec SPARK -31709 (<https://issues.apache.org/jira/browse/SPARK-31709>), cela s'applique à un plus grand nombre de cas. Par exemple, lorsque vous créez une table dans la base de données AWS Glue par défaut dont l'emplacement URI est une chaîne nulle, le message « Impossible de créer un chemin à partir d'une chaîne vide » `spark.sql("CREATE TABLE mytest (key string) location '/table_path';")` échoue. Pour contourner ce problème, définissez manuellement l'emplacement URI de vos bases de données AWS Glue, puis créez des tables dans ces bases de données à l'aide de Spark.
- Dans Amazon EMR 6.3.0, Presto SQL est passé de la version 343 à la version 350. Il y a deux changements liés à la sécurité de l'open source qui se rapportent à ce changement de version. Le contrôle d'accès au catalogue basé sur les fichiers passe de `deny` à `allow` lorsque les règles de propriété de table, de schéma ou de session ne sont pas définies. De même, le contrôle d'accès

au système basé sur les fichiers est modifié pour prendre en charge les fichiers sans règles de catalogue définies. Dans ce cas, tous les accès aux catalogues sont autorisés.

Pour plus d'informations, consultez [Version 344 \(9 octobre 2020\)](#).

- Notez que le répertoire utilisateur Hadoop (/home/hadoop) est accessible en lecture par tous. Il possède des autorisations de répertoire Unix 755 (drwxr-xr-x) pour permettre l'accès en lecture à des frameworks tels que Hive. Vous pouvez placer des fichiers dans /home/hadoop et ses sous-répertoires, mais faites attention aux autorisations sur ces répertoires pour protéger les informations sensibles.
- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux ofAmazon 2 (), qui ont un paramètre ulimit inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. Les versions EMR AMI EMR versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre ulimit EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. Le paramètre ulimit inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre ulimit lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'ulimit du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un ulimit à partir de la ligne de commande

1. Modifiez /etc/systemd/system/instance-controller.service pour ajouter les paramètres suivants à la section Service.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

Important

EMR les clusters qui exécutent Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) utilisent le comportement par défaut d'Amazon Linux et ne téléchargent ni n'installent automatiquement les mises à jour importantes et critiques du noyau nécessitant un redémarrage. Ce comportement est identique à celui des autres EC2 instances Amazon qui exécutent Amazon Linux par défaut AMI. Si de nouvelles mises à jour logicielles Amazon Linux nécessitant un redémarrage (telles que le noyau et les CUDA mises à jour) sont disponibles après la EMR sortie d'une version d'Amazon, les instances de EMR cluster qui exécutent la version par défaut AMI ne téléchargent ni n'installent automatiquement ces mises à jour. NVIDIA Pour obtenir les mises à jour du noyau,

vous pouvez [personnaliser votre Amazon EMR AMI](#) pour qu'il [utilise la dernière version d'Amazon Linux AMI](#).

- Pour utiliser les actions Spark avec Apache Oozie, vous devez ajouter la configuration suivante à votre fichier Oozie `workflow.xml`. Sinon, plusieurs bibliothèques critiques telles que Hadoop et B EMRFS seront absentes du classpath des exécuteurs Spark lancés par Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-`. Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 6.3.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.2.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.18.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.1.0	EMRConnecteur S3Select
emrfs	2,46,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.12.1	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
flink-jobmanager-config	1.12.1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-3	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.

Composant	Version	Description
hadoop-https-server	3.2.1-amzn-3	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-3	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.2.1-amzn-3	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.2.6-amzn-1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.2.6-amzn-1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.2.6-amzn-1	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	2.2.6-amzn-1	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	2.2.6-amzn-1	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	3.1.2-amzn-4	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.2-amzn-4	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	3.1.2-amzn-4	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-4	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-4	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-4	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hudi	0.7.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.7.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-prestosql	0.7.0-amzn-0	Bibliothèque de bundles pour exécuter Presto SQL avec Hudi.
hudi-spark	0.7.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.9.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.2.2	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.7.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Serveur de base de données MariaDB.

Composant	Version	Description
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5,0,0- -2,0 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	5,0,0- -2,0 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.245.1-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.245.1-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.245.1-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
prestosql-coordinator	350	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants prestosql-workers.
prestosql-worker	350	Service pour exécuter les éléments d'une requête.
prestosql-client	350	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.1.1-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.1.1-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.1.1-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.1.1-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
spark-rapids	0,4.1	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARNapplication Tez et les bibliothèques.
webserver	2.4.41+	HTTPserveur Apache.
zeppelin-server	0.9.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 6.3.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.3.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.

Classifications	Description	Actions de reconfiguration
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.

Classifications	Description	Actions de reconfiguration
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
prestosql-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoSQL)

Classifications	Description	Actions de reconfiguration
prestosql-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Modifiez les valeurs dans le fichier SQL node.properties de Presto.	Not available.
prestosql-connector-blackhole	Modifiez les valeurs dans le fichier SQL blackhole.properties de Presto.	Not available.
prestosql-connector-cassandra	Modifiez les valeurs dans le fichier SQL cassandra.properties de Presto.	Not available.
prestosql-connector-hive	Modifiez les valeurs dans le fichier SQL hive.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Modifiez les valeurs dans le fichier SQL jmx.properties de Presto.	Not available.
prestosql-connector-kafka	Modifiez les valeurs dans le fichier SQL kafka.properties de Presto.	Not available.
prestosql-connector-localfile	Modifiez les valeurs dans le fichier SQL localfile.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
prestosql-connector-memory	Modifiez les valeurs dans le fichier SQL memory.properties de Presto.	Not available.
prestosql-connector-mongodb	Modifiez les valeurs dans le fichier SQL mongodb.properties de Presto.	Not available.
prestosql-connector-mysql	Modifiez les valeurs dans le fichier SQL mysql.properties de Presto.	Not available.
prestosql-connector-postgresql	Modifiez les valeurs dans le fichier SQL postgresql.properties de Presto.	Not available.
prestosql-connector-raptor	Modifiez les valeurs dans le fichier SQL raptor.properties de Presto.	Not available.
prestosql-connector-redis	Modifiez les valeurs dans le fichier SQL redis.properties de Presto.	Not available.
prestosql-connector-redshift	Modifiez les valeurs dans le fichier SQL redshift.properties de Presto.	Not available.
prestosql-connector-tpch	Modifiez les valeurs dans le fichier SQL tpch.properties de Presto.	Not available.
prestosql-connector-tpcds	Modifiez les valeurs dans le fichier SQL tpcds.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie and HiveServer2.
yarn-env	Modifiez les valeurs dans l'YARN environnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 6.2.1

Versions d'application 6.2.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
Kit AWS SDK pour Java	1,1880	1,1880	1,1,828	1,1,828
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,10
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6-amzn-0	2.2.6-amzn-0	2.2.5	2.2.5
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,238,3	0,238,3	0.232	0.232
Spark	3.0.1	3.0.1	3.0.0	3.0.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.3.1	2.1.0	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	343	343	338	338
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 6.2.1

Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.

Modifications, améliorations et problèmes résolus

- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS

- de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
 - [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
 - [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une situation de course lors de la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
 - Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
 - Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
 - Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
 - HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPCE politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages.](#)

Problèmes connus

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement

les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :

- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 6.2.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK

Composant	Version	Description
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.0.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.16.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.0.0	EMRConnecteur S3Select
emrfs	2,44,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.11.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.11.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-2.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-2.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-2.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-2.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-2.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	3.2.1-amzn-2.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-2.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	3.2.1-amzn-2.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-2.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.2.1-amzn-2.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-2.1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.2.6-amzn-0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.2.6-amzn-0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.2.6-amzn-0	HBaseclient en ligne de commande.
hbase-rest-server	2.2.6-amzn-0	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	2.2.6-amzn-0	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	3.1.2-amzn-3	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.2-amzn-3	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-3	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	3.1.2-amzn-3	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-3	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-3	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-3	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.6.0-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.

Composant	Version	Description
hudi-presto	0.6.0-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-prestosql	0.6.0-amzn-1	Bibliothèque de bundles pour exécuter Presto SQL avec Hudi.
hudi-spark	0.6.0-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.8.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.7.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64+	Serveur de base de données MariaDB.
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5,0,0- -2,0 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	5,0,0- -2,0 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.238.3-amzn-1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.238.3-amzn-1	Service pour exécuter les éléments d'une requête.
presto-client	0.238.3-amzn-1	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
prestosql-coordinator	343	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants prestosql-workers.
prestosql-worker	343	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
prestosql-client	343	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.0.1-amzn-0.1	Clients de ligne de commande Spark.
spark-history-server	3.0.1-amzn-0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.0.1-amzn-0.1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.0.1-amzn-0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
spark-rapids	0.2.0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avec GPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	2.3.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0.9.0-preview1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 6.2.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.2.1

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classifications	Description	Actions de reconfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Not available.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Not available.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.

Classifications	Description	Actions de reconfiguration
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.

Classifications	Description	Actions de reconfiguration
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
prestosql-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoSQL)

Classifications	Description	Actions de reconfiguration
prestosql-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Modifiez les valeurs dans le fichier SQL node.properties de Presto.	Not available.
prestosql-connector-blackhole	Modifiez les valeurs dans le fichier SQL blackhole.properties de Presto.	Not available.
prestosql-connector-cassandra	Modifiez les valeurs dans le fichier SQL cassandra.properties de Presto.	Not available.
prestosql-connector-hive	Modifiez les valeurs dans le fichier SQL hive.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Modifiez les valeurs dans le fichier SQL jmx.properties de Presto.	Not available.
prestosql-connector-kafka	Modifiez les valeurs dans le fichier SQL kafka.properties de Presto.	Not available.
prestosql-connector-localfile	Modifiez les valeurs dans le fichier SQL localfile.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
prestosql-connector-memory	Modifiez les valeurs dans le fichier SQL memory.properties de Presto.	Not available.
prestosql-connector-mongodb	Modifiez les valeurs dans le fichier SQL mongodb.properties de Presto.	Not available.
prestosql-connector-mysql	Modifiez les valeurs dans le fichier SQL mysql.properties de Presto.	Not available.
prestosql-connector-postgresql	Modifiez les valeurs dans le fichier SQL postgresql.properties de Presto.	Not available.
prestosql-connector-raptor	Modifiez les valeurs dans le fichier SQL raptor.properties de Presto.	Not available.
prestosql-connector-redis	Modifiez les valeurs dans le fichier SQL redis.properties de Presto.	Not available.
prestosql-connector-redshift	Modifiez les valeurs dans le fichier SQL redshift.properties de Presto.	Not available.
prestosql-connector-tpch	Modifiez les valeurs dans le fichier SQL tpch.properties de Presto.	Not available.
prestosql-connector-tpcds	Modifiez les valeurs dans le fichier SQL tpcds.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie.
yarn-env	Modifiez les valeurs dans l'YARN environnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 6.2.0

Versions d'application 6.2.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
Kit AWS SDK pour Java	1,1880	1,1,828	1,1,828	1,1,711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,10

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.0	1.11.0	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.6-amzn-0	2.2.5	2.2.5	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.6.0-amzn-1	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1
Hue	4.8.0	4.7.1	4.7.1	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	-

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
Presto	0,238,3	0.232	0.232	0.230
Spark	3.0.1	3.0.0	3.0.0	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	-
TensorFlow	2.3.1	2.1.0	2.1.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	343	338	338	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 6.2.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.2.0 d'Amazon. Les modifications ont été apportées à la version 6.1.0.

Date de parution initiale : 9 décembre 2020

Dernière mise à jour : 4 octobre 2021

Applications prises en charge

- AWS SDK for Java version 1.11.828
- emr-record-server version 1.7.0
- Flink version 1.11.2
- Ganglia version 3.7.2
- Hadoop version 3.2.1-amzn-1
- HBaseversion 2.2.6-amzn-0
- HBase-operator-tools 1.0.0
- HCatalogversion 3.1.2-amzn-0

- Hive version 3.1.2-amzn-3
- Hudi version 0.6.0-amzn-1
- Hue version 4.8.0
- JupyterHub version 1.1.0
- Livy version 0.7.0
- MXNetversion 1.7.0
- Oozie version 5.2.0
- Phoenix version 5.0.0
- Pig version 0.17.0
- Presto version 0.238.3-amzn-1
- Presto SQL version 343
- Spark version 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow version 2.3.1
- Zeppelin version 0.9.0-preview1
- Zookeeper version 3.4.14
- Connecteurs et pilotes : Connecteur DynamoDB 4.16.0

Nouvelles fonctionnalités

- HBase: Suppression du changement de nom lors de la phase de validation et ajout d'un HFile suivi persistant. Consultez la section [HFileSuivi permanent](#) dans le guide de EMR mise à jour d'Amazon.
- HBase: Rétroporté [Créez une configuration qui oblige à mettre en cache les blocs lors du compactage](#).
- PrestoDB : améliorations apportées à l'élimination dynamique des partitions. Join Reorder basée sur des règles fonctionne sur des données non partitionnées.
- Politiques gérées délimitées : pour s'aligner sur les AWS meilleures pratiques, Amazon EMR a introduit des politiques gérées par défaut de EMR portée v2 en remplacement des politiques qui seront obsolètes. Consultez les [politiques EMR gérées par Amazon](#).
- État du support du service de métadonnées d'instance (IMDS) V2 : pour Amazon EMR 6.2 ou version ultérieure, les EMR composants Amazon sont utilisés IMDSv2 pour tous les IMDS appels. Pour les IMDS appels dans le code de votre application, vous pouvez utiliser les deux IMDSv1

etIMDSv2, ou les configurer pour qu'IMDSv1 ne soient utilisés que IMDSv2 pour renforcer la sécurité. Si vous la désactivez IMDSv1 dans les versions antérieures d'Amazon EMR 6.x, cela entraîne un échec du démarrage du cluster.

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une situation de course lors de la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/ YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Spark : amélioration des performances dans l'environnement d'exécution de Spark.

Problèmes connus

- Amazon EMR 6.2 possède des autorisations incorrectes définies sur le fichier `libinstance-controller-java /etc/cron.d/` dans la version 6.2.0. Les autorisations sur le fichier sont 645 (-rw-r--r-x), alors qu'elles devraient être 644 (-rw-r--r--). Par conséquent, la EMR version 6.2 d'Amazon n'enregistre pas les journaux d'état des instances et le répertoire `/emr/instance-logs` est vide. Ce problème est résolu dans Amazon EMR 6.3.0 et versions ultérieures.

Pour contourner ce problème, exécutez le script suivant en tant qu'action d'amorçage lors du lancement du cluster.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Pour les clusters de sous-réseaux privés Amazon EMR 6.2.0 et 6.3.0, vous ne pouvez pas accéder à l'interface utilisateur Web de Ganglia. Vous recevrez un message d'erreur « accès refusé (403) ». D'autres sites WebUIs, tels que Spark, Hue JupyterHub, Zeppelin, Livy et Tez, fonctionnent normalement. L'accès à l'interface utilisateur Web de Ganglia sur les clusters de sous-réseaux publics fonctionne également normalement. Pour résoudre ce problème, redémarrez le service `httpd` sur le nœud primaire avec `sudo systemctl restart httpd`. Ce problème est résolu dans Amazon EMR 6.4.0.
- Amazon EMR 6.2.0 présente un problème selon lequel `httpd` échoue continuellement, ce qui rend Ganglia indisponible. Le message d'erreur « Impossible de se connecter au serveur » s'affiche. Pour réparer un cluster déjà en cours d'exécution présentant ce problème, accédez SSH au nœud principal du cluster et ajoutez la ligne `Listen 80` au fichier `httpd.conf` situé dans `/etc/httpd/conf/httpd.conf`. Ce problème est résolu dans Amazon EMR 6.3.0.
- HTTPD échoue sur les clusters EMR 6.2.0 lorsque vous utilisez une configuration de sécurité. Cela rend l'interface utilisateur de l'application web Ganglia indisponible. Pour accéder à l'interface utilisateur de l'application web Ganglia, ajoutez `Listen 80` au fichier `/etc/httpd/conf/httpd.conf` sur le nœud primaire de votre cluster. Pour plus d'informations sur la connexion à votre cluster, consultez la section [Se connecter au nœud principal à l'aide SSH](#) de

EMR Les ordinateurs portables ne parviennent pas non plus à établir de connexion avec les clusters EMR 6.2.0 lorsque vous utilisez une configuration de sécurité. Le bloc-notes ne parviendra pas à répertorier les noyaux et à soumettre les tâches Spark. Nous vous recommandons d'utiliser EMR plutôt que des blocs-notes avec une autre version d'Amazon.

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux of Amazon 2 (), qui ont un paramètre `ulimit` inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre `ulimit` EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre `ulimit` inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre `ulimit` lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'`ulimit` du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un `ulimit` à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section `Service`.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer `InstanceController`

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un `ulimit` à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer `ulimit` du contrôleur d'instance à 65536 fichiers lors de la création du cluster.


```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 et 6.2.0 incluent un problème de performance qui peut affecter de manière critique toutes les opérations d'insertion, de modification et de suppression de Hudi. Si vous envisagez d'utiliser Hudi avec Amazon EMR 6.1.0 ou 6.2.0, vous devez contacter le AWS support pour obtenir un Hudi corrigé. RPM

• **⚠ Important**

EMR Les clusters qui exécutent Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) utilisent le comportement par défaut d'Amazon Linux et ne téléchargent ni n'installent automatiquement les mises à jour importantes et critiques du noyau nécessitant un redémarrage. Ce comportement est identique à celui des autres EC2 instances Amazon qui exécutent Amazon Linux par défaut AMI. Si de nouvelles mises à jour logicielles Amazon Linux nécessitant un redémarrage (telles que le noyau et les CUDA mises à jour) sont disponibles après la EMR sortie d'une version d'Amazon, les instances de EMR cluster qui exécutent la version par défaut AMI ne téléchargent ni n'installent automatiquement ces mises à jour. NVIDIA Pour obtenir les mises à jour du noyau,

vous pouvez [personnaliser votre Amazon EMR AMI](#) pour qu'il [utilise la dernière version d'Amazon Linux AMI](#).

- Les artefacts Amazon EMR 6.2.0 Maven ne sont pas publiés. Ils seront publiés avec une future version d'AmazonEMR.
- Le HFile suivi permanent à l'aide de la table système HBase Storefile ne prend pas en charge la fonctionnalité de réplification des HBase régions. Pour plus d'informations sur HBase la réplification régionale, consultez la section Nombre [élevé de lectures disponibles cohérentes avec la chronologie](#).
- Différences entre les versions de bucket d'Amazon EMR EMR 6.x et 5.x Hive

EMR La version 5.x utilise OOS Apache Hive 2, tandis que la version EMR 6.x utilise OOS Apache Hive 3. La version open source Hive2 utilise la version 1 de Bucketing, tandis que la version open source Hive3 utilise la version 2. Cette différence de version entre Hive 2 (EMR5.x) et Hive 3 (EMR6.x) signifie que le hachage des compartiments Hive fonctionne différemment. Consultez l'exemple ci-dessous.

Le tableau suivant est un exemple créé dans les versions EMR 6.x et EMR 5.x, respectivement.

```
-- Using following LOCATION in EMR 6.x
CREATE TABLE test_bucketing (id INT, desc STRING)
PARTITIONED BY (day STRING)
CLUSTERED BY(id) INTO 128 BUCKETS
LOCATION 's3://your-own-s3-bucket/emr-6-bucketing/';

-- Using following LOCATION in EMR 5.x
LOCATION 's3://your-own-s3-bucket/emr-5-bucketing/';
```

Insérer les mêmes données dans les versions EMR 6.x et EMR 5.x.

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

La vérification de l'emplacement S3 montre que le nom du fichier de compartiment est différent, car la fonction de hachage est différente entre EMR 6.x (Hive 3) et EMR 5.x (Hive 2).

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16          13 000025_0
2020-10-21 20:35:22          14 000121_0
```

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
2020-10-21 20:32:07          13 000066_0
2020-10-21 20:32:51          14 000072_0
```

Vous pouvez également voir la différence de version en exécutant la commande suivante dans Hive CLI dans la version EMR 6.x. Notez qu'il renvoie la version 2 de compartimentation.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumise. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos

paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+, -`. Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 6.2.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.0.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.16.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.0.0	EMRConnecteur S3Select
emrfs	2,44,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.11.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.11.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-2	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	3.2.1-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	3.2.1-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.2.1-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.2.6-amzn-0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.2.6-amzn-0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.2.6-amzn-0	HBaseclient en ligne de commande.

Composant	Version	Description
hbase-rest-server	2.2.6-amzn-0	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	2.2.6-amzn-0	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	3.1.2-amzn-3	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.2-amzn-3	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-3	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	3.1.2-amzn-3	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-3	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-3	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-3	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hudi	0.6.0-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.6.0-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-prestosql	0.6.0-amzn-1	Bibliothèque de bundles pour exécuter Presto SQL avec Hudi.
hudi-spark	0.6.0-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.8.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.7.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64+	Serveur de base de données MariaDB.

Composant	Version	Description
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5,0,0- -2,0 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	5,0,0- -2,0 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.238.3-amzn-1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.238.3-amzn-1	Service pour exécuter les éléments d'une requête.
presto-client	0.238.3-amzn-1	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
prestosql-coordinator	343	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants prestosql-workers.
prestosql-worker	343	Service pour exécuter les éléments d'une requête.
prestosql-client	343	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.0.1-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.0.1-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.0.1-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.0.1-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
spark-rapids	0.2.0	RAPIDSPlugin Nvidia Spark qui accélère Apache Spark avecGPUs.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.3.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARNapplication Tez et les bibliothèques.
webserver	2.4.41+	HTTPserveur Apache.
zeppelin-server	0.9.0-preview1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 6.2.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-6.2.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Not available.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Not available.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Secondary Namenode, Datanode, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	This classification should not be reconfigured.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.	Restarts Hadoop HDFS ZKFC.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat server.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier <code>hive-exec-log4j2.properties</code> de Hive.	Not available.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier <code>llap-daemon-log4j2.properties</code> de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier <code>hive-log4j2.properties</code> de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Hive	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier <code>hiveserver2-site.xml</code> de Hive Server2	Not available.

Classifications	Description	Actions de reconfiguration
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.

Classifications	Description	Actions de reconfiguration
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.

Classifications	Description	Actions de reconfiguration
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoDB)

Classifications	Description	Actions de reconfiguration
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
prestosql-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server (for PrestoSQL)

Classifications	Description	Actions de reconfiguration
prestosql-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Modifiez les valeurs dans le fichier SQL node.properties de Presto.	Not available.
prestosql-connector-blackhole	Modifiez les valeurs dans le fichier SQL blackhole.properties de Presto.	Not available.
prestosql-connector-cassandra	Modifiez les valeurs dans le fichier SQL cassandra.properties de Presto.	Not available.
prestosql-connector-hive	Modifiez les valeurs dans le fichier SQL hive.properties de Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Modifiez les valeurs dans le fichier SQL jmx.properties de Presto.	Not available.
prestosql-connector-kafka	Modifiez les valeurs dans le fichier SQL kafka.properties de Presto.	Not available.
prestosql-connector-localfile	Modifiez les valeurs dans le fichier SQL localfile.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
prestosql-connector-memory	Modifiez les valeurs dans le fichier SQL memory.properties de Presto.	Not available.
prestosql-connector-mongodb	Modifiez les valeurs dans le fichier SQL mongodb.properties de Presto.	Not available.
prestosql-connector-mysql	Modifiez les valeurs dans le fichier SQL mysql.properties de Presto.	Not available.
prestosql-connector-postgresql	Modifiez les valeurs dans le fichier SQL postgresql.properties de Presto.	Not available.
prestosql-connector-raptor	Modifiez les valeurs dans le fichier SQL raptor.properties de Presto.	Not available.
prestosql-connector-redis	Modifiez les valeurs dans le fichier SQL redis.properties de Presto.	Not available.
prestosql-connector-redshift	Modifiez les valeurs dans le fichier SQL redshift.properties de Presto.	Not available.
prestosql-connector-tpch	Modifiez les valeurs dans le fichier SQL tpch.properties de Presto.	Not available.
prestosql-connector-tpcds	Modifiez les valeurs dans le fichier SQL tpcds.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restart Oozie.
yarn-env	Modifiez les valeurs dans l'YARN environnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 6.1.1

Versions d'application 6.1.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Kit AWS SDK pour Java	1,1,828	1,1,828	1,1,711	1,1,711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,12

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 6.1.1

Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.

Modifications, améliorations et problèmes résolus

- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.

- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une situation de course lors de la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPCE politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages.](#)

Versions des composants 6.1.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-`

composant avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.14.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.0.0	EMRConnecteur S3Select
emrfs	2,42,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.11.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-1.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-1.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-1.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-1.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-1.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.2.1-amzn-1.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-1.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	3.2.1-amzn-1.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-1.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.2.1-amzn-1.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-1.1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.2.5	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.2.5	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.2.5	HBaseclient en ligne de commande.
hbase-rest-server	2.2.5	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	2.2.5	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	3.1.2-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.2-amzn-2	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-2	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	3.1.2-amzn-2	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-2	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.2-incubating-amzn-2	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.

Composant	Version	Description
hudi-presto	0.5.2-incubating-amzn-2	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Bibliothèque de bundles pour exécuter Presto SQL avec Hudi.
hudi-spark	0.5.2-incubating-amzn-2	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.7.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.6.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64+	Serveur de base de données MariaDB.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.3.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5,0,0- -2,0 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	5,0,0- -2,0 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.232	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.232	Service pour exécuter les éléments d'une requête.
presto-client	0.232	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
prestosql-coordinator	338	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants prestosql-workers.
prestosql-worker	338	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
prestosql-client	338	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.0.0-amzn-0.1	Clients de ligne de commande Spark.
spark-history-server	3.0.0-amzn-0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.0.0-amzn-0.1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	3.0.0-amzn-0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.1.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.

Composant	Version	Description
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0.9.0-preview1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 6.1.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-6.1.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.

Classifications	Description
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.

Classifications	Description
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.

Classifications	Description
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.

Classifications	Description
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.

Classifications	Description
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.

Classifications	Description
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.

Classifications	Description
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
prestoql-log	Modifiez les valeurs dans le fichier log.properties de Presto.
prestoql-config	Modifiez les valeurs dans le fichier config.properties de Presto.
prestoql-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
prestoql-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
prestoql-node	Modifiez les valeurs dans le fichier SQL node.properties de Presto.
prestoql-connector-blackhole	Modifiez les valeurs dans le fichier SQL blackhole.properties de Presto.
prestoql-connector-cassandra	Modifiez les valeurs dans le fichier SQL cassandra.properties de Presto.
prestoql-connector-hive	Modifiez les valeurs dans le fichier SQL hive.properties de Presto.
prestoql-connector-jmx	Modifiez les valeurs dans le fichier SQL jmx.properties de Presto.
prestoql-connector-kafka	Modifiez les valeurs dans le fichier SQL kafka.properties de Presto.
prestoql-connector-localfile	Modifiez les valeurs dans le fichier SQL localfile.properties de Presto.
prestoql-connector-memory	Modifiez les valeurs dans le fichier SQL memory.properties de Presto.

Classifications	Description
prestosql-connector-mongodb	Modifiez les valeurs dans le fichier SQL mongodb.properties de Presto.
prestosql-connector-mysql	Modifiez les valeurs dans le fichier SQL mysql.properties de Presto.
prestosql-connector-postgresql	Modifiez les valeurs dans le fichier SQL postgresql.properties de Presto.
prestosql-connector-raptor	Modifiez les valeurs dans le fichier SQL raptor.properties de Presto.
prestosql-connector-redis	Modifiez les valeurs dans le fichier SQL redis.properties de Presto.
prestosql-connector-redshift	Modifiez les valeurs dans le fichier SQL redshift.properties de Presto.
prestosql-connector-tpch	Modifiez les valeurs dans le fichier SQL tpch.properties de Presto.
prestosql-connector-tpcds	Modifiez les valeurs dans le fichier SQL tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Classifications	Description
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 6.1.0

Versions d'application 6.1.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Kit AWS SDK pour Java	1,1,828	1,1,828	1,1,711	1,1,711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,12

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 6.1.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.1.0 d'Amazon. Les modifications ont été apportées à la version 6.0.0.

Date de parution initiale : 4 septembre 2020

Dernière mise à jour : 15 octobre 2020

Applications prises en charge

- AWS SDK for Java version 1.11.828
- Flink version 1.11.0
- Ganglia version 3.7.2
- Hadoop version 3.2.1-amzn-1
- HBaseversion 2.2.5
- HBase-operator-tools 1.0.0
- HCatalogversion 3.1.2-amzn-0
- Hive version 3.1.2-amzn-1

- Hudi version 0.5.2-incubation
- Hue version 4.7.1
- JupyterHub version 1.1.0
- Livy version 0.7.0
- MXNetversion 1.6.0
- Oozie version 5.2.0
- Phoenix version 5.0.0
- Presto version 0.232
- Presto SQL version 338
- Spark version 3.0.0-amzn-0
- TensorFlow version 2.1.0
- Zeppelin version 0.9.0-preview1
- Zookeeper version 3.4.14
- Connecteurs et pilotes : Connecteur DynamoDB 4.14.0

Nouvelles fonctionnalités

- ARM Les types d'instance sont pris en charge à partir de la EMR version 5.30.0 d'Amazon et de la EMR version 6.1.0 d'Amazon.
- Les types d'instances à usage général m6G sont pris en charge à partir des EMR versions 6.1.0 et 5.30.0 d'Amazon. Pour plus d'informations, consultez la section [Types d'instances pris en charge](#) dans le guide EMR de gestion Amazon.
- La fonctionnalité de groupe de EC2 placement est prise en charge à partir de EMR la version 5.23.0 d'Amazon en tant qu'option pour plusieurs clusters de nœuds principaux. Actuellement, seuls les types de nœuds primaires sont pris en charge par la fonctionnalité de groupe de placement. La stratégie SPREAD est appliquée à ces nœuds primaires. La stratégie SPREAD place un petit groupe d'instances sur un matériel sous-jacent distinct afin de se prémunir contre la perte de plusieurs nœuds primaires en cas de panne matérielle. Pour plus d'informations, consultez la section [EMRIntégration au groupe de EC2 placement](#) dans le guide EMR de gestion Amazon.
- Dimensionnement géré : avec Amazon EMR version 6.1.0, vous pouvez activer le dimensionnement EMR géré par Amazon pour augmenter ou diminuer automatiquement le nombre d'instances ou d'unités dans votre cluster en fonction de la charge de travail. Amazon évalue EMR en permanence les indicateurs des clusters afin de prendre des décisions de dimensionnement

qui optimisent les coûts et la rapidité de vos clusters. Managed Scaling est également disponible sur Amazon EMR version 5.30.0 et versions ultérieures, à l'exception de la version 6.0.0. Pour plus d'informations, consultez la section [Scaling Cluster Resources](#) dans le guide EMR de gestion Amazon.

- La SQL version 338 de Presto est prise en charge avec la version EMR 6.1.0. Pour plus d'informations, consultez [Presto](#).
 - Presto SQL est pris en charge sur les versions EMR 6.1.0 et ultérieures uniquement, pas sur les versions EMR 6.0.0 ou 5.x. EMR
 - Le nom de l'application, Presto continue d'être utilisé pour installer PrestoDB sur des clusters. Pour installer Presto SQL sur des clusters, utilisez le nom PrestoSQL de l'application.
 - Vous pouvez installer PrestoDB ou SQL Presto, mais vous ne pouvez pas installer les deux sur un seul cluster. Si PrestoDB et SQL Presto sont spécifiés lors de la tentative de création d'un cluster, une erreur de validation se produit et la demande de création de cluster échoue.
 - Presto SQL est pris en charge à la fois sur les clusters à maître unique et à maîtres multiples. Sur les clusters multi-maîtres, un métastore Hive externe est requis pour exécuter Presto ou PrestoDB. SQL Consultez la section [Applications prises en charge dans un EMR cluster comportant plusieurs nœuds principaux](#).
- ECR prise en charge de l'authentification automatique sur Apache Hadoop et Apache Spark avec Docker : les utilisateurs de Spark peuvent utiliser des images Docker provenant de Docker Hub et d'Amazon Elastic Container Registry (ECRAmazon) pour définir les dépendances de l'environnement et des bibliothèques.

[Configurez Docker](#) et [exécutez des applications Spark avec Docker à l'aide d'Amazon EMR 6.x](#).

- EMR prend en charge les ACID transactions Apache Hive : Amazon EMR 6.1.0 ajoute la prise en charge des ACID transactions Hive afin qu'elles soient conformes aux ACID propriétés d'une base de données. Grâce à cette fonctionnalité, vous pouvez exécuter les opérations INSERT, UPDATE, DELETE, et MERGE dans des tables gérées par Hive avec des données stockées dans Amazon Simple Storage Service (Amazon S3). Il s'agit d'une fonctionnalité clé pour les cas d'utilisation tels que l'ingestion du streaming, le retraitement des données MERGE, l'utilisation de mises à jour groupées et la modification lente des dimensions. Pour plus d'informations, notamment des exemples de configuration et des cas d'utilisation, consultez [Amazon EMR supporte les ACID transactions Apache Hive](#).

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une situation de course lors de la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Apache Flink n'est pas pris en charge sur la EMR version 6.0.0, mais il l'est sur la version EMR 6.1.0 avec Flink 1.11.0. Il s'agit de la première version de Flink officiellement compatible avec Hadoop 3. Consultez [Annonce de sortie d'Apache Flink 1.11.0](#).
- Ganglia a été supprimé des ensembles de packages EMR 6.1.0 par défaut.

Problèmes connus

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux ofAmazon 2 (), qui ont un paramètre ulimit inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre ulimit EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre ulimit inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre ulimit lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'ulimit du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un ulimit à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section Service.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

• **⚠ Important**

Amazon EMR 6.1.0 et 6.2.0 incluent un problème de performance qui peut affecter de manière critique toutes les opérations d'insertion, de modification et de suppression de Hudi. Si vous envisagez d'utiliser Hudi avec Amazon EMR 6.1.0 ou 6.2.0, vous devez contacter le AWS support pour obtenir un Hudi corrigé. RPM

- Si vous définissez une configuration personnalisée de collecte des déchets avec `spark.driver.extraJavaOptions` et `spark.executor.extraJavaOptions`, cela entraînera un échec du lancement du pilote/exécuteur avec la EMR version 6.1 en raison d'une configuration de collecte des déchets conflictuelle. Avec la EMR version 6.1.0, vous devez spécifier une configuration personnalisée de collecte des déchets Spark pour les pilotes et les exécuteurs avec les propriétés `spark.driver.defaultJavaOptions` et `spark.executor.defaultJavaOptions` à la place. Pour en savoir plus, consultez [l'environnement d'exécution Apache Spark](#) et [la configuration de Spark Garbage Collection sur Amazon EMR 6.1.0](#).

- L'utilisation de Pig avec Oozie (et au sein de Hue, puisque Hue utilise les actions Oozie pour exécuter les scripts Pig), génère une erreur indiquant qu'une bibliothèque native-lzo ne peut pas être chargée. Ce message d'erreur est informatif et n'empêche pas Pig de fonctionner.
- Prise en charge de la simultanéité dans Hudi : Actuellement, Hudi ne prend pas en charge les écritures simultanées dans une seule table Hudi. De plus, Hudi annule toutes les modifications effectuées par les enregistreurs en cours avant de permettre à un nouvel enregistreur de démarrer. Les écritures simultanées peuvent interférer avec ce mécanisme et introduire des conditions de concurrence, ce qui peut entraîner une corruption des données. Vous devez vous assurer que, dans le cadre de votre flux de traitement des données, il n'y a qu'un seul enregistreur Hudi opérant sur une table Hudi à tout moment. Hudi prend en charge plusieurs lecteurs simultanés opérant sur la même table Hudi.
- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos

paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Amazon EMR 6.1.0 présente un problème qui affecte les clusters exécutant Presto. Après une période prolongée (plusieurs jours), le cluster peut générer des erreurs telles que "su : failed to execute /bin/bash : Resource temporarily unavailable" ou "shell request failed on channel 0". Ce problème est dû à un EMR processus interne d'Amazon (InstanceController) qui génère trop de processus légers (LWP), ce qui finit par amener l'utilisateur Hadoop à dépasser sa limite nproc. Cela empêche l'utilisateur d'ouvrir des processus supplémentaires. La solution à ce problème consiste à effectuer une mise à niveau vers la version EMR 6.2.0.

Versions des composants 6.1.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-goodies	3.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.14.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	2.0.0	EMRConnecteur S3Select
emrfs	2,42,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.11.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	3.2.1-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.2.1-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	3.2.1-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.2.5	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.2.5	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	2.2.5	HBaseclient en ligne de commande.
hbase-rest-server	2.2.5	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.2.5	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	3.1.2-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	3.1.2-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-2	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.2-amzn-2	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-2	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.2-incubating-amzn-2	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.5.2-incubating-amzn-2	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hudi-prestosql	0.5.2-incubating-amzn-2	Bibliothèque de bundles pour exécuter Presto SQL avec Hudi.

Composant	Version	Description
hudi-spark	0.5.2-incubating-amzn-2	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hue-server	4.7.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.6.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64+	Serveur de base de données MariaDB.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.3.0	Bibliothèque open source Vision par ordinateur

Composant	Version	Description
phoenix-library	5,0,0- -2,0 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	5,0,0- -2,0 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.232	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.232	Service pour exécuter les éléments d'une requête.
presto-client	0.232	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
prestoql-coordinator	338	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants prestoql-workers.
prestoql-worker	338	Service pour exécuter les éléments d'une requête.
prestoql-client	338	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	2.0.0	Système de gestion des clés Apache Ranger
spark-client	3.0.0-amzn-0	Clients de ligne de commande Spark.
spark-history-server	3.0.0-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	3.0.0-amzn-0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	3.0.0-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.1.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.

Composant	Version	Description
zeppelin-server	0.9.0-preview1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 6.1.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-6.1.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.

Classifications	Description
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.

Classifications	Description
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.

Classifications	Description
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.

Classifications	Description
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier <code>oozie-log4j.properties</code> d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier <code>oozie-site.xml</code> d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.

Classifications	Description
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.

Classifications	Description
prestosql-log	Modifiez les valeurs dans le fichier log.properties de Presto.
prestosql-config	Modifiez les valeurs dans le fichier config.properties de Presto.
prestosql-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
prestosql-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
prestosql-node	Modifiez les valeurs dans le fichier SQL node.properties de Presto.
prestosql-connector-blackhole	Modifiez les valeurs dans le fichier SQL blackhole.properties de Presto.
prestosql-connector-cassandra	Modifiez les valeurs dans le fichier SQL cassandra.properties de Presto.
prestosql-connector-hive	Modifiez les valeurs dans le fichier SQL hive.properties de Presto.
prestosql-connector-jmx	Modifiez les valeurs dans le fichier SQL jmx.properties de Presto.
prestosql-connector-kafka	Modifiez les valeurs dans le fichier SQL kafka.properties de Presto.
prestosql-connector-localfile	Modifiez les valeurs dans le fichier SQL localfile.properties de Presto.
prestosql-connector-memory	Modifiez les valeurs dans le fichier SQL memory.properties de Presto.
prestosql-connector-mongodb	Modifiez les valeurs dans le fichier SQL mongodb.properties de Presto.

Classifications	Description
prestosql-connector-mysql	Modifiez les valeurs dans le fichier SQL mysql.properties de Presto.
prestosql-connector-postgresql	Modifiez les valeurs dans le fichier SQL postgresql.properties de Presto.
prestosql-connector-raptor	Modifiez les valeurs dans le fichier SQL raptor.properties de Presto.
prestosql-connector-redis	Modifiez les valeurs dans le fichier SQL redis.properties de Presto.
prestosql-connector-redshift	Modifiez les valeurs dans le fichier SQL redshift.properties de Presto.
prestosql-connector-tpch	Modifiez les valeurs dans le fichier SQL tpch.properties de Presto.
prestosql-connector-tpcds	Modifiez les valeurs dans le fichier SQL tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 6.0.1

Versions d'application 6.0.1

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Kit AWS SDK pour Java	1,1,828	1,1,828	1,1,711	1,1,711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,12
AmazonCloudWatchAgent	-	-	-	-

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 6.0.1

Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.

Modifications, améliorations et problèmes résolus

- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une situation de course lors de la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.

- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPC politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages.](#)

Versions des composants 6.0.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.14.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.5.0	EMRConnecteur S3Select
emrfs	2.39.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par

Composant	Version	Description
		le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-0.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-0.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-0.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-0.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-0.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	3.2.1-amzn-0.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-0.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-0.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-0.1	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	3.2.1-amzn-0.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-0.1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.2.3	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.2.3	HBaseclient en ligne de commande.
hbase-rest-server	2.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.2.3	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	3.1.2-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.2-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.

Composant	Version	Description
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	3.1.2-amzn-0	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-0	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.0-incubating-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.5.0-incubating-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.0.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	REST interface pour interagir avec Apache Spark

Composant	Version	Description
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.5.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64+	Serveur de base de données MariaDB.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5,0,0- -2,0 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	5,0,0- -2,0 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.230	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.230	Service pour exécuter les éléments d'une requête.
presto-client	0.230	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
r	3.4.3	Projet R pour les calculs statistiques
spark-client	2.4.4	Clients de ligne de commande Spark.
spark-history-server	2.4.4	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.4	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.4.4	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.

Composant	Version	Description
zeppelin-server	0,9,0- SNAPSHOT	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 6.0.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-6.0.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.

Classifications	Description
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-PROTO-HIVE-SITE	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive

Classifications	Description
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.

Classifications	Description
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 6.0.0

Versions d'application 6.0.0

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Kit AWS SDK pour Java	1,1,828	1,1,828	1,1,711	1,1,711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,10	2,1,10	2,1,10	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	2.2.5	2.2.5	2.2.3	2.2.3

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
HCatalog	3.1.2	3.1.2	3.1.2	3.1.2
Hadoop	3.2.1	3.2.1	3.2.1	3.2.1
Hive	3.1.2	3.1.2	3.1.2	3.1.2
Hudi	0.5.2-incubating-amzn-2	0.5.2-incubating-amzn-2	0.5.0-incubating-amzn-1	0.5.0-incubating-amzn-1
Hue	4.7.1	4.7.1	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	-	-	-	-
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	-	-
Presto	0.232	0.232	0.230	0.230
Spark	3.0.0	3.0.0	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	-	-
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Trino (PrestoSQL)	338	338	-	-
Zeppelin	0.9.0	0.9.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 6.0.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.0.0 d'Amazon.

Date de parution initiale : 10 mars 2020

Applications prises en charge

- AWS SDK for Java version 1.11.711
- Ganglia version 3.7.2
- Hadoop version 3.2.1
- HBaseversion 2.2.3
- HCatalogversion 3.1.2
- Hive version 3.1.2
- Hudi version 0.5.0-incubation
- Hue version 4.4.0
- JupyterHub version 1.0.0
- Livy version 0.6.0
- MXNetversion 1.5.1
- Oozie version 5.1.0
- Phoenix version 5.0.0
- Presto version 0.230
- Spark version 2.4.4

- TensorFlow version 1.14.0
- Zeppelin version 0.9.0- SNAPSHOT
- Zookeeper version 3.4.14
- Connecteurs et pilotes : Connecteur DynamoDB 4.14.0

Note

Flink, Sqoop, Pig et Mahout ne sont pas disponibles dans la version 6.0.0 d'Amazon. EMR

Nouvelles fonctionnalités

- YARN Docker Runtime Support : YARN les applications, telles que les tâches Spark, peuvent désormais s'exécuter dans le contexte d'un conteneur Docker. Cela vous permet de définir facilement les dépendances dans une image Docker sans avoir à installer de bibliothèques personnalisées sur votre EMR cluster Amazon. Pour plus d'informations, consultez [Configurer l'intégration de Docker](#) et [exécuter des applications Spark avec Docker à l'aide d'Amazon EMR 6.0.0](#).
- LLAP Support Hive - Hive prend désormais en charge le mode LLAP d'exécution pour améliorer les performances des requêtes. Pour plus d'informations, consultez la section [Utilisation de Hive. LLAP](#)

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.

- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une situation de course lors de la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Amazon Linux
 - Amazon Linux 2 est le système d'exploitation de la série de versions EMR 6.x.
 - `systemd` utilisé pour la gestion des services au lieu de `upstart` in Amazon Linux 1.
- Kit de développement Java (JDK)
 - JDK Corretto 8 est la version JDK par défaut pour la série de versions 6.x EMR.
- Scala
 - Scala 2.12 est utilisé avec Apache Spark et Apache Livy.
- Python 3
 - Python 3 est désormais la version par défaut de Python dans EMR.
- YARN Étiquettes de nœuds
 - À compter de la série de versions Amazon EMR 6.x, la fonctionnalité d'étiquettes de YARN nœuds est désactivée par défaut. Les processus principaux des applications peuvent s'exécuter à la fois sur les nœuds de noyau et sur les nœuds de tâche par défaut. Vous pouvez activer la fonctionnalité d'étiquettes de YARN nœuds en configurant les propriétés suivantes : `yarn.node-labels.enabled` et `yarn.node-labels.am.default-node-label-expression`. Pour plus d'informations, consultez [Comprendre les nœuds primaires, les nœuds principaux et les nœuds de tâches](#).

Problèmes connus

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux of Amazon 2 (), qui ont un paramètre ulimit inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre ulimit EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre ulimit inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre ulimit lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'ulimit du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un ulimit à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section Service.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Le shell interactif Spark PySpark, y compris SparkR et spark-shell, ne prend pas en charge l'utilisation de Docker avec des bibliothèques supplémentaires.
- Pour utiliser Python 3 avec Amazon EMR version 6.0.0, vous devez ajouter PATH à `yarn.nodemanager.env-whitelist`.
- Les fonctionnalités Live Long et Process (LLAP) ne sont pas prises en charge lorsque vous utilisez le catalogue de données AWS Glue comme métastore pour Hive.
- Lorsque vous utilisez Amazon EMR 6.0.0 avec l'intégration de Spark et Docker, vous devez configurer les instances de votre cluster avec le même type d'instance et le même nombre de EBS volumes afin d'éviter toute défaillance lors de la soumission d'une tâche Spark avec Docker Runtime.
- Dans Amazon EMR 6.0.0, HBase le mode de stockage Amazon S3 est concerné par le problème [HBASE-24286](#). HBasemaster ne peut pas s'initialiser lorsque le cluster est créé à l'aide de données S3 existantes.
- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 6.0.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	3.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.14.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.5.0	EMRConnecteur S3Select
emrfs	2.39.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les

Composant	Version	Description
		métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	3.2.1-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	3.2.1-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	3.2.1-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	3.2.1-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	3.2.1-amzn-0	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	3.2.1-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	3.2.1-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	3.2.1-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	3.2.1-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	3.2.1-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	3.2.1-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	2.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	2.2.3	Service pour desservir une ou plusieurs HBase régions.
hbase-client	2.2.3	HBaseclient en ligne de commande.
hbase-rest-server	2.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	2.2.3	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hcatalog-client	3.1.2-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	3.1.2-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	3.1.2-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	3.1.2-amzn-0	Client de ligne de commande Hive.
hive-hbase	3.1.2-amzn-0	Client Hive-hbase.
hive-metastore-server	3.1.2-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	3.1.2-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.0-incubating-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.5.0-incubating-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.

Composant	Version	Description
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.0.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mxnet	1.5.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64+	Serveur de base de données MariaDB.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	5,0,0- -2,0 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	5,0,0- -2,0 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.230	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.230	Service pour exécuter les éléments d'une requête.
presto-client	0.230	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
r	3.4.3	Projet R pour les calculs statistiques
spark-client	2.4.4	Clients de ligne de commande Spark.
spark-history-server	2.4.4	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.4	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.4	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.41+	HTTP serveur Apache.
zeppelin-server	0,9,0- SNAPSHOT	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 6.0.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-6.0.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.

Classifications	Description
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.

Classifications	Description
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-env	Modifiez les valeurs dans l'HDFS environnement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive	Paramètres EMR sélectionnés par Amazon pour Apache Hive.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.

Classifications	Description
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.

Classifications	Description
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier <code>oozie-log4j.properties</code> d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier <code>oozie-site.xml</code> d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.

Classifications	Description
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.

Classifications	Description
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS

Classifications	Description
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Versions publiées par Amazon EMR 5.x

Cette section contient les versions des applications, les notes de mise à jour, les versions des composants et les classifications de configuration disponibles dans chaque version d'Amazon EMR 5.x.

Lorsque vous lancez un cluster, vous pouvez choisir parmi plusieurs versions d'AmazonEMR. Vous pouvez ainsi tester et utiliser des versions d'applications qui répondent à vos exigences de compatibilité. Vous spécifiez le numéro de version avec l'étiquette de version. Les étiquettes de version sont sous la forme `emr-x.x.x`. Par exemple, `emr-7.2.0`.

Les nouvelles EMR versions d'Amazon sont mises à disposition dans différentes régions sur une période de plusieurs jours, en commençant par la première région à la date de sortie initiale. Il est possible que la dernière version ne soit pas disponible dans votre région pendant cette période.

Pour un tableau complet des versions des applications dans chaque version d'Amazon EMR 5.x, consultez [Versions des applications dans les versions 5.x d'Amazon EMR](#).

Rubriques

- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Amazon EMR version 5.36.2](#)
- [Amazon EMR version 5.36.1](#)
- [Amazon EMR version 5.36.0](#)
- [Amazon EMR version 5.35.0](#)
- [Amazon EMR version 5.34.0](#)

- [Amazon EMR version 5.33.1](#)
- [Amazon EMR version 5.33.0](#)
- [Amazon EMR version 5.32.1](#)
- [Amazon EMR version 5.32.0](#)
- [Amazon EMR version 5.31.1](#)
- [Amazon EMR version 5.31.0](#)
- [Amazon EMR version 5.30.2](#)
- [Amazon EMR version 5.30.1](#)
- [Amazon EMR version 5.30.0](#)
- [Amazon EMR version 5.29.0](#)
- [Amazon EMR version 5.28.1](#)
- [Amazon EMR version 5.28.0](#)
- [Amazon EMR version 5.27.1](#)
- [Amazon EMR version 5.27.0](#)
- [Amazon EMR version 5.26.0](#)
- [Amazon EMR version 5.25.0](#)
- [Amazon EMR version 5.24.1](#)
- [Amazon EMR version 5.24.0](#)
- [Amazon EMR version 5.23.1](#)
- [Amazon EMR version 5.23.0](#)
- [Amazon EMR version 5.22.0](#)
- [Amazon EMR version 5.21.2](#)
- [Amazon EMR version 5.21.1](#)
- [Amazon EMR version 5.21.0](#)
- [Amazon EMR version 5.20.1](#)
- [Amazon EMR version 5.20.0](#)
- [Amazon EMR version 5.19.1](#)
- [Amazon EMR version 5.19.0](#)
- [Amazon EMR version 5.18.1](#)

- [Amazon EMR version 5.18.0](#)
- [Amazon EMR version 5.17.2](#)
- [Amazon EMR version 5.17.1](#)
- [Amazon EMR version 5.17.0](#)
- [Amazon EMR version 5.16.1](#)
- [Amazon EMR version 5.16.0](#)
- [Amazon EMR version 5.15.1](#)
- [Amazon EMR version 5.15.0](#)
- [Amazon EMR version 5.14.2](#)
- [Amazon EMR version 5.14.1](#)
- [Amazon EMR version 5.14.0](#)
- [Amazon EMR version 5.13.1](#)
- [Amazon EMR version 5.13.0](#)
- [Amazon EMR version 5.12.3](#)
- [Amazon EMR version 5.12.2](#)
- [Amazon EMR version 5.12.1](#)
- [Amazon EMR version 5.12.0](#)
- [Amazon EMR version 5.11.4](#)
- [Amazon EMR version 5.11.3](#)
- [Amazon EMR version 5.11.2](#)
- [Amazon EMR version 5.11.1](#)
- [Amazon EMR version 5.11.0](#)
- [Amazon EMR version 5.10.1](#)
- [Amazon EMR version 5.10.0](#)
- [Amazon EMR version 5.9.1](#)
- [Amazon EMR version 5.9.0](#)
- [Amazon EMR version 5.8.3](#)
- [Amazon EMR version 5.8.2](#)
- [Amazon EMR version 5.8.1](#)

- [Amazon EMR version 5.8.0](#)
- [Amazon EMR version 5.7.1](#)
- [Amazon EMR version 5.7.0](#)
- [Amazon EMR version 5.6.1](#)
- [Amazon EMR version 5.6.0](#)
- [Amazon EMR version 5.5.4](#)
- [Amazon EMR version 5.5.3](#)
- [Amazon EMR version 5.5.2](#)
- [Amazon EMR version 5.5.1](#)
- [Amazon EMR version 5.5.0](#)
- [Amazon EMR version 5.4.1](#)
- [Amazon EMR version 5.4.0](#)
- [Amazon EMR version 5.3.2](#)
- [Amazon EMR version 5.3.1](#)
- [Amazon EMR version 5.3.0](#)
- [Amazon EMR version 5.2.3](#)
- [Amazon EMR version 5.2.2](#)
- [Amazon EMR version 5.2.1](#)
- [Amazon EMR version 5.2.0](#)
- [Amazon EMR version 5.1.1](#)
- [Amazon EMR version 5.1.0](#)
- [Amazon EMR version 5.0.3](#)
- [Amazon EMR version 5.0.2](#)
- [Amazon EMR version 5.0.1](#)
- [Amazon EMR version 5.0.0](#)

Versions des applications dans les versions 5.x d'Amazon EMR

Pour obtenir un tableau complet répertoriant les versions d'applications disponibles dans chaque version d'Amazon EMR 5.x, ouvrez [Versions des applications dans les versions 5.x d'Amazon EMR](#) dans votre navigateur.

Amazon EMR version 5.36.2

versions de l'application 5.36.2

Les applications suivantes sont prises en charge dans cette version : [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
Kit AWS SDK pour Java	1,12,206	1,12,206	1,12,206	1,12,159
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.2	1.14.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
HCatalog	2.3.9	2.3.9	2.3.9	2.3.9
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.9	2.3.9
Hudi	0.10.1-amzn-1	0.10.1-amzn-1	0.10.1-amzn-1	0.9.0-amzn-2
Hue	4.10.0	4.10.0	4.10.0	4.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.6.0	2.6.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,267	0,267	0,267	0,266
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.10.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

notes de mise à jour de la version 5.36.2

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.36.2 d'Amazon. Les modifications sont relatives à la version 5.36.1. Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Modifications, améliorations et problèmes résolus

- Cette version améliore la logique de réduction du cluster afin qu'Amazon EMR ne réduise pas les nœuds principaux en dessous du paramètre de facteur de HDFS réplication défini pour le cluster. Cette amélioration répond aux exigences de redondance des données et réduit le risque de blocage d'une opération de dimensionnement.
- Cette version ajoute un nouveau mécanisme de nouvelle tentative au flux de travail de dimensionnement des clusters pour ceux qui exécutent Presto ou Trino. Cette amélioration réduit le risque que le redimensionnement du cluster s'exécute indéfiniment en raison de l'échec d'une seule opération de redimensionnement. Cela améliore également l'utilisation du cluster, car celui-ci augmente et diminue la capacité plus rapidement.
- Résout un problème selon lequel les opérations de réduction du cluster pouvaient être bloquées pendant qu'Amazon mettait hors service EMR gracieusement un nœud principal et que celui-ci devenait inutilisable avant sa mise hors service complète.
- Améliore la stabilité d'un nœud dans un cluster à haute disponibilité comportant plusieurs nœuds principaux lorsqu'Amazon EMR redémarre un seul nœud.
- Optimise la gestion des journaux avec Amazon EMR exécuté sur AmazonEC2. C'est pourquoi vous constaterez peut-être une légère réduction des coûts de stockage pour les journaux de votre cluster.

- Améliore la gestion des fichiers journaux de ZooKeeper transactions conservés sur les nœuds principaux afin de minimiser les scénarios dans lesquels les fichiers journaux dépassent les limites et interrompent les opérations du cluster.
- Corrige un bogue rare qui peut provoquer l'échec d'un cluster à haute disponibilité comportant plusieurs nœuds principaux en raison de l'impossibilité de communiquer avec le Yarn ResourceManager.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 503.0	4,14,343	xxxxxx, 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

Versions des composants 5.36.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon

EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.16.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-notebook-env</code>	1.5.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
<code>emr-s3-dist-cp</code>	2.21.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.7.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2,51,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.14.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.14.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.1-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.1-amzn-4	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	2.10.1-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.1-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.9-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.9-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.9-amzn-2	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.9-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.9-amzn-2	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.9-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.9-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.10.1-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.10.1-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.10.1-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.13.1	nginx [engine x] est HTTP un serveur proxy inverse

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.8.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68	Mon serveur SQL de base de données.
nvidia-cuda	11,0,194	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.267-amzn-1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.267-amzn-1	Service pour exécuter les éléments d'une requête.
presto-client	0.267-amzn-1	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.8-amzn-2	Clients de ligne de commande Spark.
spark-history-server	2.4.8-amzn-2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.8-amzn-2	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.4.8-amzn-2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.10.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

5.36.2 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

classifications EMR-5.36.2

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classifications	Description	Actions de reconfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	Should not be reconfigured.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Actions de reconfiguration
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Classifications	Description	Actions de reconfiguration
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.

Classifications	Description	Actions de reconfiguration
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.	Restarts Presto-Server.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.	Restarts Presto-Server.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.	Restarts EMR record server.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.	Restarts EMR record server.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.	Restarts EMR record server.
spark	Paramètres EMR sélection nés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restarts Oozie.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

5.36.2 journal des modifications

Journal des modifications pour la version 5.36.2 et notes de version

Date	Événement	Description
28/05/2024	Déploiement terminé	Amazon EMR 5.36.2 entièrement déployé dans toutes les régions prises en charge
28/05/2024	Publication de documents	Notes de mise à jour d'Amazon EMR 5.36.2 publiées pour la première fois
20/05/2024/	Première version	Amazon EMR 5.36.2 a été déployé pour la première fois dans des régions commerciales limitées

Amazon EMR version 5.36.1

Versions d'application 5.36.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
Kit AWS SDK pour Java	1,12,206	1,12,206	1,12,159	1,1970
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.14.2	1.13.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.9	2.3.9	2.3.9	2.3.8
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.9	2.3.8
Hudi	0.10.1-amzn-1	0.10.1-amzn-1	0.9.0-amzn-2	0.9.0-amzn-0

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
Hue	4.10.0	4.10.0	4.10.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.6.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.4.1
Livy	0.7.1	0.7.1	0.7.1	0.7.1
MXNet	1.8.0	1.8.0	1.8.0	1.8.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.1
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,267	0,267	0,266	0,261
Spark	2.4.8	2.4.8	2.4.8	2.4.8
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.10.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.36.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.36.1 d'Amazon. Les modifications ont été apportées à la version 5.36.0. Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Modifications, améliorations et problèmes résolus

- La EMR version 5.36.1 d'Amazon ajoute la prise en charge de l'archivage des journaux dans Amazon S3 lors de la réduction du cluster. Dans les versions 5.x précédentes, vous pouviez uniquement archiver les fichiers journaux sur Amazon S3 lors de la résiliation du cluster. Cette amélioration garantit que les fichiers journaux générés sur le cluster sont conservés sur Amazon S3 même après la résiliation du nœud. Pour plus d'informations, consultez [Configuration de la journalisation et du débogage de cluster](#).
- La version 5.36.1 améliore le démon de gestion des journaux sur le cluster afin de surveiller des dossiers de journaux supplémentaires dans votre cluster. EMR Cette amélioration permet de minimiser les scénarios de surutilisation des disques.
- La version 5.36.1 redémarre automatiquement le démon de gestion des journaux du cluster lorsqu'il s'arrête. Cette amélioration réduit le risque que les nœuds apparaissent défectueux en raison d'une surutilisation du disque.
- La version 5.36.1 corrige un problème selon lequel les EMR démons Amazon sur le nœud principal conservaient des métadonnées périmées pour les instances terminées du cluster. Le maintien de données périmées peut entraîner une augmentation illimitée de l'utilisation du cluster CPU et de la mémoire, ce qui peut entraîner des défaillances du cluster.
- Pour les clusters lancés avec plusieurs nœuds principaux, la version 5.36.1 résout un problème selon lequel une défaillance EC2 matérielle d'Amazon sur l'un des nœuds principaux pouvait provoquer la défaillance d'un deuxième nœud principal et rendre votre cluster instable.
- Pour les clusters configurés avec un chiffrement en transit, Managed Scaling prend désormais en charge les données de réorganisation Spark. Les données de réorganisation Spark sont des données que Spark redistribue entre les partitions pour effectuer des opérations spécifiques. Lors de la réduction, Managed Scaling ignore les instances contenant des données de réorganisation. Cela permet d'éviter les nouvelles tentatives de tâches et les nouveaux calculs, qui sont coûteux en termes de prix et de performances. Pour plus d'informations sur les opérations de réorganisation, consultez le [Guide de programmation Spark](#).
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou

Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023.0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

Versions des composants 5.36.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version

d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.16.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.5.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.21.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.7.0	EMRConnecteur S3Select
emrfs	2,51,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
flink-client	1.14.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.14.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.1-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.1-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httfs-server	2.10.1-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.1-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.

Composant	Version	Description
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.9-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.9-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.9-amzn-2	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.9-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.9-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.9-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server2	2.3.9-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.10.1-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.10.1-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.10.1-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.13.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.8.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.

Composant	Version	Description
mariadb-server	5,5,68	Mon serveur SQL de base de données.
nvidia-cuda	11,0,194	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.267-amzn-1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.267-amzn-1	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
presto-client	0.267-amzn-1	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.8-amzn-2	Clients de ligne de commande Spark.
spark-history-server	2.4.8-amzn-2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.8-amzn-2	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.4.8-amzn-2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.

Composant	Version	Description
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.10.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.36.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-5.36.1

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the ResourceManager service.

Classifications	Description	Actions de reconfiguration
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	Should not be reconfigured.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Actions de reconfiguration
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Classifications	Description	Actions de reconfiguration
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
hudi-defaults	Modifiez les valeurs dans le fichier hudi-defaults.conf de Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.

Classifications	Description	Actions de reconfiguration
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.	Restarts Presto-Server.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.	Restarts Presto-Server.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.

Classifications	Description	Actions de reconfiguration
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.	Restarts EMR record server.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.	Restarts EMR record server.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.	Restarts EMR record server.
spark	Paramètres EMR sélection nés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restarts Oozie.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Journal des modifications 5.36.1

Journal des modifications pour la version 5.36.1 et les notes de version

Date	Événement	Description
26/07/2023	Mettre à jour	Étiquettes 2.0.20230612.0 et 2.0.20230628.0 de la nouvelle version du système d'exploitation.
25/05/2023	Déploiement terminé	Amazon EMR 5.36.1 entièrement déployé dans toutes les régions prises en charge
09/05/2023	Publication de documents	Notes de mise à jour d'Amazon EMR 5.36.1 publiées pour la première fois
04/05/2023	Première version	Amazon EMR 5.36.1 a été déployé pour la première fois dans des régions commerciales limitées

Amazon EMR version 5.36.0

Versions d'application 5.36.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
Kit AWS SDK pour Java	1,12,206	1,12,159	1,1970	1,1970
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.14.2	1.13.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
HCatalog	2.3.9	2.3.9	2.3.8	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.9	2.3.8	2.3.7
Hudi	0.10.1-amzn-1	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1
Hue	4.10.0	4.10.0	4.9.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.4.1	1.2.2
Livy	0.7.1	0.7.1	0.7.1	0.7.0
MXNet	1.8.0	1.8.0	1.8.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.1	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,267	0,266	0,261	0,245,1
Spark	2.4.8	2.4.8	2.4.8	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.10.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.36.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.36.0 d'Amazon. Les modifications ont été apportées à la version 5.35.0.

Date de parution initiale : 15 juin 2022

Nouvelles fonctions

- La EMR version 5.36.0 d'Amazon ajoute la prise en charge du langage de définition des données (DDL) avec Apache Spark sur les clusters compatibles avec Apache Ranger. Cela vous permet d'utiliser Apache Ranger pour gérer l'accès à des opérations telles que la création, la modification et la suppression de bases de données et de tables d'un EMR cluster Amazon.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

Note

Cette version ne reçoit plus de AMI mises à jour automatiques puisqu'elle a été remplacée par une autre version de correctif. La version du correctif est indiquée par le numéro qui suit la deuxième décimale (6.8.1). Pour savoir si vous utilisez la dernière version du correctif, consultez les versions disponibles dans le [guide des versions](#), ou consultez le menu déroulant des EMRversions d'Amazon lorsque vous créez un cluster dans la console, ou utilisez l'[list-release-labels](#) CLI action [ListReleaseLabels](#) API ou. Pour être tenu au courant des nouvelles versions, abonnez-vous au RSS fil d'actualité sur le [site Quoi de neuf ?](#) page.

OsRelea Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv)

OsRelea Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelea Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres) , Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong- Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Paci fique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 307.0	4,14,305	15 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 207.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023119.1	4,14,301	3 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202210.1	4,14,301	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 103.3	4,14.296	5 décembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelea Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 004.0	4,14.294	2 novembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres) , Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong- Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Paci fique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 912.1	4,14,291	7 octobre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0,2022 719,0	4,14.287	10 août 2022	USA Ouest (Californie du Nord), Europe (Paris), Europe (Stockholm), Europe (Francfort), Asie-Pacifique (Mumbai), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 426.0	4,14,281	14 juin 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

Modifications, améliorations et problèmes résolus

- Les mises à niveau EMR vers Amazon 5.36.0 sont désormais prises en charge : aws-sdk 1.12.206, Hadoop 2.10.1-amzn-4, Hive 2.3.9-amzn-2, Hudi 0.10.1-amzn-1, Spark 2.4.8-amzn-2, Presto 0.267-amzn-1, connecteur Amazon Glue 1.18.0, 2.51.0. EMRFS

Problèmes connus

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

- Avec les EMR versions 5.36.0 et 6.6.0 à 6.9.0 d'Amazon, `SecretAgent` les composants du `RecordServer` service peuvent subir une perte de données de journal en raison d'une configuration incorrecte du modèle de nom de fichier dans les propriétés `Log4j2`. En cas de configuration incorrecte, les composants ne génèrent qu'un seul fichier journal par jour. Lorsque la stratégie de rotation est appliquée, elle remplace le fichier existant au lieu de générer un nouveau fichier journal comme prévu. Pour contourner le problème, utilisez une action d'amorçage pour générer des journaux toutes les heures et ajoutez un nombre entier auto-incrémenté dans le nom du fichier pour gérer la rotation.

Pour les versions Amazon EMR 6.6.0 à 6.9.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Pour Amazon EMR 5.36.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Versions des composants 5.36.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.16.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-notebook-env	1.5.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.21.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.7.0	EMRConnecteur S3Select
emrfs	2,51,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.14.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.14.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.10.1-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.1-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.10.1-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.1-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.9-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.9-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.

Composant	Version	Description
hcatalog-webhcat-server	2.3.9-amzn-2	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.9-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.9-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.9-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.9-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.10.1-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.10.1-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.10.1-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.13.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.8.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68	Mon serveur SQL de base de données.
nvidia-cuda	11,0,194	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.267-amzn-1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.267-amzn-1	Service pour exécuter les éléments d'une requête.
presto-client	0.267-amzn-1	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.8-amzn-2	Clients de ligne de commande Spark.
spark-history-server	2.4.8-amzn-2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	2.4.8-amzn-2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.8-amzn-2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.10.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.36.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-5.36.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the Resource Manager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop

Classifications	Description	Actions de reconfiguration
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	Should not be reconfigured.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Actions de reconfiguration
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Classifications	Description	Actions de reconfiguration
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.

Classifications	Description	Actions de reconfiguration
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server.

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier <code>redis.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.	Restarts EMR record server.

Classifications	Description	Actions de reconfiguration
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier <code>erver.properties</code> .	Restarts EMR record server.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier <code>log4j.properties</code> .	Restarts EMR record server.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier <code>spark-defaults.conf</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier <code>metrics.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier <code>oraoop-site.xml</code> de OraOop Sqoop.	Not available.

Classifications	Description	Actions de reconfiguration
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restarts Oozie.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 5.35.0

Versions d'application 5.35.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
Kit AWS SDK pour Java	1,12,159	1,1970	1,1970	1,1970
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.14.2	1.13.1	1.12.1	1.12.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
HCatalog	2.3.9	2.3.8	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.9	2.3.8	2.3.7	2.3.7
Hudi	0.9.0-amzn-2	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1
Hue	4.10.0	4.9.0	4.9.0	4.9.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.4.1	1.2.2	1.2.2
Livy	0.7.1	0.7.1	0.7.0	0.7.0
MXNet	1.8.0	1.8.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.1	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,266	0,261	0,245,1	0,245,1
Spark	2.4.8	2.4.8	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.4.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.10.0	0.9.0	0.9.0
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.35.0

Ceci est la note de mise à jour de la EMR version 5.35.0 d'Amazon.

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.35.0 d'Amazon. Les modifications ont été apportées à la version 5.34.0.

Date de parution initiale : 30 mars 2022

Nouvelles fonctions

- Les applications Amazon EMR version 5.35 qui utilisent Log4j 1.x et Log4j 2.x sont mises à niveau pour utiliser Log4j 1.2.17 (ou version ultérieure) et Log4j 2.17.1 (ou version ultérieure) respectivement, et ne nécessitent pas d'actions d'amorçage pour atténuer les problèmes rencontrés dans les versions précédentes. CVE veuillez consulter [Approche d'atténuation CVE-2021-44228](#).

Modifications, améliorations et problèmes résolus

Modifications apportées à Flink

Type de modification	Description
Mises à niveau	<ul style="list-style-type: none"> Mise à jour de la version de flink vers la version 1.14.2. log4j a été mis à jour vers la version 2.17.1.

Modifications apportées à Hadoop

Type de modification	Description
Rétroportages open source Hadoop depuis la version 5.34.0 EMR	<ul style="list-style-type: none"> • YARN-10438 : Gérer la valeur nulle container Id en C <code>ClientRMService # getContainerReport ()</code> • YARN-7266 : Les fils du gestionnaire d'événements de Timeline Server sont verrouillés • YARN-10438 : ATS 1.5 ne démarre pas si RollingLevelDb les fichiers sont corrompus ou manquants • HADOOP-13500 : Synchronisation de l'itération de l'objet des propriétés de configuration • YARN-10651 : CapacityScheduler s'est écrasé avec NPE in. <code>AbstractYarnScheduler updateNodeResource()</code> • HDFS-12221 : Remplacer les xerces dans <code>XmlEditsVisitor</code> • HDFS-16410 : Analyse XML non sécurisée dans <code>OfflineEditsXMLLoader</code>
Modifications et corrections apportées à Hadoop	<ul style="list-style-type: none"> • Tomcat est utilisé dans KMS et HTTPFS est mis à niveau vers la version 8.5.75 • Dans la <code>FileSystemOptimizedCommitter</code> version 2, le marqueur de réussite était écrit dans le chemin <code>commitJob</code> de sortie défini lors de la création du commit. Étant donné que <code>commitJob</code> les chemins de sortie au niveau des tâches peuvent différer, le chemin a été corrigé pour utiliser celui défini dans les fichiers manifestes. Pour les tâches Hive, cela signifie que le marqueur de réussite est correctement écrit lors de l'exécution

Type de modification	Description
	d'opérations telles que la partition dynamique ou UNIONALL.

Modifications apportées à Hive

Type de modification	Description
Hive est passé à la version open source 2.3.9 , y compris ces correctifs JIRA	<ul style="list-style-type: none"> • HIVE-17155 : findConfFile () dans le HiveConf fichier .java présente quelques problèmes avec le chemin de configuration • HIVE-24797 : Désactiver la validation des valeurs par défaut lors de l'analyse des schémas Avro • HIVE-21563 : Améliorez les getEmptyTable performances de Table# en désactivant Once registerAllFunctions • HIVE-18147 : Les tests peuvent échouer avec java.net. BindException: Adresse déjà utilisée • HIVE-24608 : Revenir à get_table dans HMS le client pour Hive 2.3.x • HIVE-21200 : Vectorisation : colonne de date lançant java.lang. UnsupportedOperationException pour parquet • HIVE-19228 : Supprime l'utilisation de commons-httpclient 3.x
Rétroportages open source Hive depuis la version 5.34.0 EMR	<ul style="list-style-type: none"> • HIVE-19990 : La requête avec un intervalle littéral dans la condition de jointure échoue • HIVE-25824 : Mise à niveau de la branche 2.3 vers log4j 2.17.0

Type de modification	Description
	<ul style="list-style-type: none"> • TEZ-4062 : La planification des tentatives spéculatives doit être abandonnée une fois la tâche terminée • TEZ-4108 : NullPointerException lors d'une exécution spéculative, condition de course • TEZ-3918 : La configuration de <code>tez.task.log.level</code> ne fonctionne pas
Mises à niveau et correctifs apportés à Hive	<ul style="list-style-type: none"> • Mise à niveau de la version de Log4j vers 2.17.1 • Mise à niveau de ORC la version vers la version 1.4.3 • Correction d'un blocage dû à un fil de pénalité ShuffleScheduler
Nouvelles fonctionnalités	<ul style="list-style-type: none"> • Ajout d'une fonctionnalité permettant d'imprimer Hive Query dans les journaux AM. Par défaut, l'option est désactivée. Drapeau/Conf : <code>tez.am.emr.print.hive.query.in.log</code> État (par défaut) :FALSE.

Modifications apportées à Oozie

Type de modification	Description
Rétroportages open source d'Oozie depuis la version 5.34.0 EMR	<ul style="list-style-type: none"> • OOZIE-3652 : Le lanceur Oozie devrait réessayer de répertorier les répertoires lorsque cela se produit NoSuchFileException

Modifications apportées à Pig

Type de modification	Description
Mises à niveau	<ul style="list-style-type: none">log4j a été mis à jour vers la version 1.2.17.

Problèmes connus

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.35.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.15.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.5.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.20.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.7.0	EMRConnecteur S3Select
emrfs	2,49,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.14.2	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
flink-jobmanager-config	1.14.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.1-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.1-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.1-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.10.1-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.1-amzn-3	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.

Composant	Version	Description
hadoop-https-server	2.10.1-amzn-3	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.1-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-3	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.1-amzn-3	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.1-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.9-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.9-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.9-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.9-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.9-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.9-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.9-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hudi	0.9.0-amzn-2	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.9.0-amzn-2	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.9.0-amzn-2	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.10.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.13.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.8.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68	Mon serveur SQL de base de données.

Composant	Version	Description
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.266-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.266-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.266-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.8-amzn-1	Clients de ligne de commande Spark.
spark-history-server	2.4.8-amzn-1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.8-amzn-1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.8-amzn-1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0.10.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.35.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-5.35.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the Resource Manager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	Should not be reconfigured.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Actions de reconfiguration
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Classifications	Description	Actions de reconfiguration
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.

Classifications	Description	Actions de reconfiguration
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server.

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier <code>redis.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.	Restarts EMR record server.

Classifications	Description	Actions de reconfiguration
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier <code>erver.properties</code> .	Restarts EMR record server.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier <code>log4j.properties</code> .	Restarts EMR record server.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier <code>spark-defaults.conf</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier <code>metrics.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier <code>oraoop-site.xml</code> de OraOop Sqoop.	Not available.

Classifications	Description	Actions de reconfiguration
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restarts Oozie.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 5.34.0

Versions d'application 5.34.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
Kit AWS SDK pour Java	1,1970	1,1970	1,1970	1,1,890
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.13.1	1.12.1	1.12.1	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
HCatalog	2.3.8	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.8	2.3.7	2.3.7	2.3.7
Hudi	0.9.0-amzn-0	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0
Hue	4.9.0	4.9.0	4.9.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.4.1	1.2.2	1.2.2	1.1.0
Livy	0.7.1	0.7.0	0.7.0	0.7.0
MXNet	1.8.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.1	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,261	0,245,1	0,245,1	0,240,1
Spark	2.4.8	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.4.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.10.0	0.9.0	0.9.0	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.34.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.34.0 d'Amazon. Les modifications ont été apportées à la version 5.33.1.

Date de parution initiale : 20 janvier 2022

Date de mise à niveau : 21 mars 2022

Nouvelles fonctions

- [Dimensionnement géré] Optimisation du dimensionnement géré par les données Spark Shuffle - Pour EMR les versions 5.34.0 et ultérieures d'Amazon, ainsi que pour EMR les versions 6.4.0 et ultérieures, le dimensionnement géré prend désormais en compte les données Spark shuffle (données que Spark redistribue entre les partitions pour effectuer des opérations spécifiques). Pour plus d'informations sur les opérations de shuffle, consultez la section [Utilisation du dimensionnement EMR géré EMR dans Amazon](#) dans le Guide de EMRgestion Amazon et le Guide de [programmation Spark](#).
- [Hudi] Améliorations visant à simplifier la configuration de Hudi. Le contrôle de simultanéité optimiste a été désactivé par défaut.

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Auparavant, le redémarrage manuel du gestionnaire de ressources sur un cluster multi-maîtres obligeait les EMR démons Amazon du cluster, tels que Zookeeper, à recharger tous les nœuds précédemment désaffectés ou perdus dans le fichier znode de Zookeeper. Cela a entraîné le

dépassement des limites par défaut dans certaines situations. Amazon supprime EMR désormais les enregistrements de nœuds mis hors service ou perdus datant de plus d'une heure du fichier Zookeeper et les limites internes ont été augmentées.

- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état du nœud, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Mise à niveau de Zeppelin vers la version 0.10.0.
- Livy Fix – mise à niveau vers la version 0.7.1
- Amélioration des performances de Spark : les exécuteurs hétérogènes sont désactivés lorsque certaines valeurs de configuration de Spark sont remplacées dans la version 5.34.0. EMR
- Le Web HDFS et le serveur HTTPFS sont désactivés par défaut. Vous pouvez réactiver le Web à HDFS l'aide de la configuration Hadoop, `dfs.webhdfs.enabled` Le serveur HttpFS peut être démarré en utilisant `sudo systemctl start hadoop-httpfs`.

Problèmes connus

- La fonctionnalité Amazon EMR Notebooks utilisée avec l'usurpation d'identité d'un utilisateur de Livy ne fonctionne pas car le protocole HTTPFS est désactivé par défaut. Dans ce cas, le EMR bloc-notes ne peut pas se connecter au cluster sur lequel l'emprunt d'identité Livy est activé. La solution consiste à démarrer le serveur HTTPFS avant de connecter le EMR bloc-notes au cluster à l'aide de `sudo systemctl start hadoop-httpfs`
- Les requêtes Hue ne fonctionnent pas dans Amazon EMR 6.4.0 car le serveur Apache Hadoop HTTPFS est désactivé par défaut. Pour utiliser Hue on Amazon EMR 6.4.0, démarrez manuellement le serveur HTTPFS sur le nœud EMR principal Amazon à l'aide `sudo systemctl start hadoop-httpfs` ou [utilisez une étape Amazon EMR](#).
- La fonctionnalité Amazon EMR Notebooks utilisée avec l'usurpation d'identité d'un utilisateur de Livy ne fonctionne pas car le protocole HTTPFS est désactivé par défaut. Dans ce cas, le EMR bloc-notes ne peut pas se connecter au cluster sur lequel l'emprunt d'identité Livy est activé. La solution consiste à démarrer le serveur HTTPFS avant de connecter le EMR bloc-notes au cluster à l'aide de `sudo systemctl start hadoop-httpfs`
- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.34.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.14.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-notebook-env</code>	1.4.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
<code>emr-s3-dist-cp</code>	2.18.0	Application de copie distribuée optimisée pour Amazon S3.

Composant	Version	Description
emr-s3-select	1.7.0	EMRConnecteur S3Select
emrfs	2,48,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.13.1	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.13.1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.1-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.1-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.10.1-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.10.1-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.1-amzn-2	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.10.1-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.1-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.1-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.1-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.8-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.8-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.8-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.8-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.8-amzn-0	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.8-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.8-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.9.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.9.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.9.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.9.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.4.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.1-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.13.1	nginx [engine x] est HTTP un serveur proxy inverse

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.8.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68	Mon serveur SQL de base de données.
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.1	Client de ligne de commande Oozie.
oozie-server	5.2.1	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.261-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.261-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.261-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.8-amzn-0	Clients de ligne de commande Spark.
spark-history-server	2.4.8-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.8-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.4.8-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.10.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.34.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-5.34.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classifications	Description	Actions de reconfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	Should not be reconfigured.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Actions de reconfiguration
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Classifications	Description	Actions de reconfiguration
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.

Classifications	Description	Actions de reconfiguration
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server.

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier <code>redis.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.	Restarts EMR record server.

Classifications	Description	Actions de reconfiguration
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier <code>erver.properties</code> .	Restarts EMR record server.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier <code>log4j.properties</code> .	Restarts EMR record server.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier <code>spark-defaults.conf</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier <code>metrics.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier <code>oraoop-site.xml</code> de OraOop Sqoop.	Not available.

Classifications	Description	Actions de reconfiguration
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restarts Oozie.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 5.33.1

Versions d'application 5.33.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
Kit AWS SDK pour Java	1,1970	1,1970	1,1,890	1,1,890
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.12.1	1.11.2	1.11.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.1
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.7.0-amzn-1	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.9.0	4.9.0	4.8.0	4.8.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	2.1.0
JupyterHub	1.2.2	1.2.2	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,245,1	0,245,1	0,240,1	0,240,1
Spark	2.4.7	2.4.7	2.4.7	2.4.7
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.4.1	2.3.1	2.3.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.9.0	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.33.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.33.0/5.33.1 d'Amazon. Les modifications ont été apportées à la version 5.32.0.

Date de parution initiale : 19 avril 2021

Dernière mise à jour : 9 août 2021

Mises à niveau

- Mise à niveau du connecteur Amazon Glue vers la version 1.15.0
- Mise à niveau AWS SDK for Java vers la version 1.11.970
- Mise à niveau EMRFS vers la version 2.46.0
- EMRGoodies mis à jour vers la version 2.14.0
- Serveur d'EMRenregistrement mis à niveau vers la version 1.9.0
- EMRS3 Dist CP a été mis à niveau vers la version 2.18.0
- Agent EMR secret mis à niveau vers la version 1.8.0
- Mise à niveau de Flink vers la version 1.12.1
- Mise à niveau de Hadoop vers la version 2.10.1-amzn-1
- Mise à niveau de Hive vers la version 2.3.7-amzn-4
- Mise à niveau de Hudi vers la version 0.7.0
- Mise à niveau de Hue vers la version 4.9.0
- Mise à niveau de OpenCV vers la version 4.5.0
- Mise à niveau de Presto vers la version 0.245.1-amzn-0
- Mise à niveau de R vers la version 4.0.2

- Mise à niveau de Spark vers la version 2.4.7-amzn-1
- Mise à niveau TensorFlow vers la version 2.4.1
- Mise à niveau de Zeppelin vers la version 0.9.0

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état du nœud, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite maximale de fichiers ouverts sur les anciennes versions AL2 d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.

- Configuration d'un cluster pour résoudre les problèmes de performances du serveur Apache YARN Timeline versions 1 et 1.5

Les versions 1 et 1.5 d'Apache YARN Timeline Server peuvent entraîner des problèmes de performances avec de grands EMR clusters très actifs, en particulier avec `yarn.resourcemanager.system-metrics-publisher.enabled=true`, qui est le paramètre par défaut d'AmazonEMR. Un YARN Timeline Server v2 open source résout le problème de performance lié à l'évolutivité du YARN Timeline Server.

Les autres solutions à ce problème incluent :

- Configuration de `yarn.resourcemanager.system-metrics-publisher.enabled=false` dans le fichier `yarn-site.xml`.
- Activation du correctif pour ce problème lors de la création d'un cluster, comme décrit ci-dessous.

Les EMR versions Amazon suivantes contiennent un correctif pour ce problème de performance du serveur YARN Timeline.

EMR5,30.2, 5,31.1, 5,32.1, 5,33.1, 5,34.x, 6,0.1, 6.1.1, 6.2.1, 6.3.1, 6.4.x

Pour activer le correctif sur l'une des EMR versions Amazon spécifiées ci-dessus, définissez ces propriétés sur `true` dans un JSON fichier de configuration transmis à l'aide du [paramètre de aws emr create-cluster commande](#) :`--configurations file://./configurations.json`. Vous pouvez également activer le correctif à l'aide de l'[interface utilisateur de la console de reconfiguration](#).

Exemple du contenu du fichier `configurations.json` :

```
[
{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
      "true",
    "yarn.resourcemanager.system-metrics-publisher.enabled": "true"
  },
  "Configurations": []
}
```

- L'exécution de Spark est maintenant plus rapide lors de la récupération des emplacements de partition à partir de Hive Metastore pour les requêtes d'insertion de Spark.
- Versions de composants mises à niveau. Pour obtenir la liste des versions des composants, consultez la section [À propos des EMR versions d'Amazon](#) dans ce guide.
- Le SDK bundle AWS Java a été installé sur chaque nouveau cluster. Il s'agit d'un fichier jar unique contenant tous les services SDKs et leurs dépendances, au lieu de fichiers jar de composants individuels. Pour plus d'informations, consultez [Java SDK Bundled Dependency](#).
- Les problèmes de Managed Scaling rencontrés dans les EMR versions précédentes d'Amazon ont été corrigés et des améliorations ont été apportées afin de réduire de manière significative les taux d'échec des applications.
- HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPC politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages.](#)

Nouvelles fonctionnalités

- Amazon EMR prend en charge les points d'accès Amazon S3, une fonctionnalité d'Amazon S3 qui vous permet de gérer facilement l'accès aux lacs de données partagés. En utilisant votre alias Amazon S3 Access Point, vous pouvez simplifier l'accès aux données à grande échelle sur AmazonEMR. Vous pouvez utiliser les points d'accès Amazon S3 avec toutes les versions d'Amazon EMR sans frais supplémentaires dans toutes les AWS régions où Amazon EMR est disponible. Pour en savoir plus sur les points d'accès Amazon S3 et les alias de points d'accès, consultez [Utilisation d'un alias de type compartiment pour votre point d'accès](#) dans le Guide de l'utilisateur Amazon S3.
- Amazon EMR -5.33 prend en charge les nouveaux types d'EC2 instances Amazon : c5a, c5ad, c6gn, c6gd, m6gd, d3, d3en, m5zn, r5b, r6gd. Consultez [Types d'instances pris en charge](#).

Problèmes connus

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0,

emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux ofAmazon 2 (), qui ont un paramètre `ulimit` inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par défaut. Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre `ulimit` par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. Le paramètre `ulimit` inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre `ulimit` lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'`ulimit` du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un `ulimit` à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section `Service`.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer `InstanceController`

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```


Définissez un `ulimit` à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer `ulimit` du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
```

```
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Pour les clusters de sous-réseaux privés Amazon EMR 6.3.0 et 6.2.0, vous ne pouvez pas accéder à l'interface utilisateur Web de Ganglia. Vous recevrez un message d'erreur « accès refusé (403) ». D'autres sites WebUIs, tels que Spark, Hue JupyterHub, Zeppelin, Livy et Tez, fonctionnent normalement. L'accès à l'interface utilisateur Web de Ganglia sur les clusters de sous-réseaux publics fonctionne également normalement. Pour résoudre ce problème, redémarrez le service httpd sur le nœud primaire avec `sudo systemctl restart httpd`. Ce problème est résolu dans Amazon EMR 6.4.0.

 Important

EMR les clusters qui exécutent Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) utilisent le comportement par défaut d'Amazon Linux et ne téléchargent ni n'installent automatiquement les mises à jour importantes et critiques du noyau nécessitant un redémarrage. Ce comportement est identique à celui des autres EC2 instances Amazon qui exécutent Amazon Linux par défaut AMI. Si de nouvelles mises à jour logicielles Amazon Linux nécessitant un redémarrage (telles que le noyau et les CUDA mises à jour) sont disponibles après la EMR sortie d'une version d'Amazon, les instances de EMR cluster qui exécutent la version par défaut AMI ne téléchargent ni n'installent automatiquement ces mises à jour. NVIDIA Pour obtenir les mises à jour du noyau, vous pouvez [personnaliser votre Amazon EMR AMI](#) pour qu'il [utilise la dernière version d'Amazon Linux AMI](#).

- La prise en charge par console pour créer une configuration de sécurité spécifiant l'option d'intégration de AWS Ranger n'est actuellement pas prise en charge dans la GovCloud région. La configuration de sécurité peut être effectuée à l'aide du CLI. Consultez la section [Création EMR de la configuration de sécurité](#) dans le guide EMR de gestion Amazon.
- Politiques gérées délimitées : pour s'aligner sur les AWS meilleures pratiques, Amazon EMR a introduit des politiques gérées par défaut EMR définies dans la version v2 en remplacement des politiques qui seront déconseillées. Consultez les [politiques EMR gérées par Amazon](#).
- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.33.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.14.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.2.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.18.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.6.0	EMRConnecteur S3Select
emrfs	2,46,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.12.1	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
flink-jobmanager-config	1.12.1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.1-amzn-1.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.1-amzn-1.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.1-amzn-1.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.10.1-amzn-1.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.1-amzn-1.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.

Composant	Version	Description
hadoop-https-server	2.10.1-amzn-1.1	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.1-amzn-1.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-1.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-1.1	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.1-amzn-1.1	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.1-amzn-1.1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.7-amzn-4	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.7-amzn-4	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.7-amzn-4	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.7-amzn-4	Client de ligne de commande Hive.
hive-hbase	2.3.7-amzn-4	Client Hive-hbase.
hive-metastore-server	2.3.7-amzn-4	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.7-amzn-4	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hudi	0.7.0-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.7.0-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.7.0-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.9.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.2.2	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.7.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 ans et plus	Mon serveur SQL de base de données.

Composant	Version	Description
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.245.1-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.245.1-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.245.1-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.7-amzn-1.1	Clients de ligne de commande Spark.
spark-history-server	2.4.7-amzn-1.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.7-amzn-1.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.7-amzn-1.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0.9.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.33.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-5.33.1

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.

Classifications	Description	Actions de reconfiguration
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/Yarn.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, HiveMetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	Should not be reconfigured.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Actions de reconfiguration
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Classifications	Description	Actions de reconfiguration
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.

Classifications	Description	Actions de reconfiguration
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server.

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier <code>redis.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.	Restarts EMR record server.

Classifications	Description	Actions de reconfiguration
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier <code>erver.properties</code> .	Restarts EMR record server.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier <code>log4j.properties</code> .	Restarts EMR record server.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier <code>spark-defaults.conf</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier <code>metrics.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier <code>oraoop-site.xml</code> de OraOop Sqoop.	Not available.

Classifications	Description	Actions de reconfiguration
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restarts Oozie.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 5.33.0

Versions d'application 5.33.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
Kit AWS SDK pour Java	1,1970	1,1,890	1,1,890	1,1,852
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.12.1	1.11.2	1.11.2	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.1	2.10.0
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.7.0-amzn-1	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.9.0	4.8.0	4.8.0	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	2.1.0	-
JupyterHub	1.2.2	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.7.0	1.6.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,245,1	0,240,1	0,240,1	0,238,3
Spark	2.4.7	2.4.7	2.4.7	2.4.6
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.4.1	2.3.1	2.3.1	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.9.0	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.33.0

Versions des composants 5.33.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.14.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.2.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.18.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.6.0	EMRConnecteur S3Select
emrfs	2,46,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.12.1	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.12.1	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.1-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.1-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.1-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.10.1-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.1-amzn-1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.10.1-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.1-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.10.1-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.1-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.1-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.7-amzn-4	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.7-amzn-4	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.7-amzn-4	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.7-amzn-4	Client de ligne de commande Hive.
hive-hbase	2.3.7-amzn-4	Client Hive-hbase.
hive-metastore-server	2.3.7-amzn-4	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.7-amzn-4	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.7.0-amzn-1	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.7.0-amzn-1	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.7.0-amzn-1	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.

Composant	Version	Description
hue-server	4.9.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.2.2	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.7.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68	Mon serveur SQL de base de données.
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.5.0	Bibliothèque open source Vision par ordinateur

Composant	Version	Description
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.245.1-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.245.1-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.245.1-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	4.0.2	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.7-amzn-1	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.4.7-amzn-1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.7-amzn-1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.7-amzn-1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.4.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.9.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.33.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-5.33.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Classifications	Description	Actions de reconfiguration
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegistrator, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
flink-log4j-session	Modifiez les paramètres de Flink log4j-session.properties pour la session Kubernetes/ Yarn.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.

Classifications	Description	Actions de reconfiguration
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.
hbase	Paramètres EMR sélectionnés par Amazon pour Apache HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	Should not be reconfigured.

Classifications	Description	Actions de reconfiguration
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.

Classifications	Description	Actions de reconfiguration
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.

Classifications	Description	Actions de reconfiguration
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.

Classifications	Description	Actions de reconfiguration
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l'MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.

Classifications	Description	Actions de reconfiguration
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.	Not available.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.

Classifications	Description	Actions de reconfiguration
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.	Restarts EMR record server.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.	Restarts EMR record server.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.	Restarts EMR record server.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.	Restarts Spark history server and Spark thrift server.

Classifications	Description	Actions de reconfiguration
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.	Not available.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restarts Oozie.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.

Classifications	Description	Actions de reconfiguration
zeppelin-site	Modification des paramètres de configuration dans zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 5.32.1

Versions d'application 5.32.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
Kit AWS SDK pour Java	1,1,890	1,1,890	1,1,852	1,1,852
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.2	1.11.0	1.11.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.7	2.3.7	2.3.7
Hadoop	2.10.1	2.10.1	2.10.0	2.10.0
Hive	2.3.7	2.3.7	2.3.7	2.3.7
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0
Hue	4.8.0	4.8.0	4.7.1	4.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	2.1.0	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.7.0	1.6.0	1.6.0

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,240,1	0,240,1	0,238,3	0,238,3
Spark	2.4.7	2.4.7	2.4.6	2.4.6
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.3.1	2.1.0	2.1.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.32.1

Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.

Modifications, améliorations et problèmes résolus

- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état du nœud, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas

en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR

- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite maximale de fichiers ouverts sur les anciennes versions AL2 d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPCE politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages.](#)

Problèmes connus

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement

les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :

- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+, -`. Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.32.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK

Composant	Version	Description
emr-ddb	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.13.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.1.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
emr-s3-dist-cp	2.17.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.6.0	EMRConnecteur S3Select
emrfs	2,45,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.11.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.11.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.1-amzn-0.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.1-amzn-0.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.1-amzn-0.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.10.1-amzn-0.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.1-amzn-0.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	2.10.1-amzn-0.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.1-amzn-0.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.10.1-amzn-0.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-0.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.1-amzn-0.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.1-amzn-0.1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hcatalog-client	2.3.7-amzn-3	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.7-amzn-3	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.7-amzn-3	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.7-amzn-3	Client de ligne de commande Hive.
hive-hbase	2.3.7-amzn-3	Client Hive-hbase.
hive-metastore-server	2.3.7-amzn-3	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.7-amzn-3	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.6.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.6.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.

Composant	Version	Description
hudi-presto	0.6.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.8.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.7.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68 et plus	Mon serveur SQL de base de données.
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
opencv	4.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.240.1-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.240.1-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.240.1-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.7-amzn-0.1	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.4.7-amzn-0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.7-amzn-0.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.7-amzn-0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.3.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.32.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-5.32.1

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier <code>container-executor.cfg</code> YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Classifications	Description	Actions de reconfiguration
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.	Not available.

Classifications	Description	Actions de reconfiguration
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	Should not be reconfigured.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Actions de reconfiguration
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Classifications	Description	Actions de reconfiguration
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.

Classifications	Description	Actions de reconfiguration
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server.

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier <code>redis.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.	Restarts EMR record server.

Classifications	Description	Actions de reconfiguration
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier <code>erver.properties</code> .	Restarts EMR record server.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier <code>log4j.properties</code> .	Restarts EMR record server.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier <code>spark-defaults.conf</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier <code>metrics.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier <code>oraoop-site.xml</code> de OraOop Sqoop.	Not available.

Classifications	Description	Actions de reconfiguration
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restarts Oozie.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 5.32.0

Versions d'application 5.32.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
Kit AWS SDK pour Java	1,1,890	1,1,852	1,1,852	1,1,759
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.2	1.11.0	1.11.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
HCatalog	2.3.7	2.3.7	2.3.7	2.3.6
Hadoop	2.10.1	2.10.0	2.10.0	2.8.5
Hive	2.3.7	2.3.7	2.3.7	2.3.6
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.6.0-amzn-0	0.5.2-incubating
Hue	4.8.0	4.7.1	4.7.1	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	2.1.0	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.7.0	1.6.0	1.6.0	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,240,1	0,238,3	0,238,3	0.232
Spark	2.4.7	2.4.6	2.4.6	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.3.1	2.1.0	2.1.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.32.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.32.0 d'Amazon. Les modifications ont été apportées à la version 5.31.0.

Date de parution initiale : 8 janvier 2021

Mises à niveau

- Mise à niveau du connecteur Amazon Glue vers la version 1.14.0
- Mise à niveau d'Amazon SageMaker Spark SDK vers la version 1.4.1
- Mise à niveau AWS SDK for Java vers la version 1.11.890
- Connecteur EMR DynamoDB version 4.16.0 mise à niveau
- Mise à niveau EMRFS vers la version 2.45.0
- Mise à niveau de EMR Log Analytics Metrics vers la version 1.18.0
- EMR MetricsAndEventsApiGateway Client mis à niveau vers la version 1.5.0
- Serveur d'EMRenregistrement mis à niveau vers la version 1.8.0
- Mise à niveau de EMR S3 Dist CP vers la version 2.17.0
- Agent EMR secret mis à niveau vers la version 1.7.0
- Mise à niveau de Flink vers la version 1.11.2
- Mise à niveau de Hadoop vers la version 2.10.1-amzn-0
- Mise à niveau de Hive vers la version 2.3.7-amzn-3
- Mise à niveau de Hue vers la version 4.8.0
- Mise à niveau de Mxnet vers la version 1.7.0
- Mise à niveau d'OpenCV vers la version 4.4.0

- Mise à niveau de Presto vers la version 0.240.1-amzn-0
- Mise à niveau de Spark vers la version 2.4.7-amzn-0
- Mise à niveau TensorFlow vers la version 2.3.1

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état du nœud, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite maximale de fichiers ouverts sur les anciennes versions AL2 d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.

- Versions de composants mises à niveau.
- Pour obtenir la liste des versions des composants, consultez la section [À propos des EMR versions d'Amazon](#) dans ce guide.

Nouvelles fonctionnalités

- À partir d'Amazon EMR 5.32.0 et 6.5.0, le dimensionnement dynamique de l'exécuteur pour Apache Spark est activé par défaut. Pour activer ou désactiver cette fonctionnalité, vous pouvez utiliser le paramètre de configuration `spark.yarn.heterogeneousExecutors.enabled`.
- État de prise en charge du service de métadonnées d'instance (IMDS) V2 : les composants Amazon EMR 5.23.1, 5.27.1 et 5.32 ou versions ultérieures sont utilisés pour tous les appels. IMDSv2 IMDS Pour les IMDS appels dans le code de votre application, vous pouvez utiliser les deux IMDSv1 et IMDSv2, ou les configurer pour qu'IMDSv1 ne soient utilisés que IMDSv2 pour renforcer la sécurité. Pour les autres EMR versions 5.x, la désactivation IMDSv1 entraîne l'échec du démarrage du cluster.
- À partir d'Amazon EMR 5.32.0, vous pouvez lancer un cluster qui s'intègre nativement à Apache Ranger. Apache Ranger est un cadre open source permettant d'activer, de surveiller et de gérer la sécurité globale des données sur la plateforme Hadoop. Pour plus d'informations, consultez [Apache Ranger](#). Grâce à l'intégration native, vous pouvez utiliser votre propre Apache Ranger pour appliquer un contrôle précis de l'accès aux données sur Amazon. EMR Consultez la section [Intégrer Amazon EMR à Apache Ranger](#) dans le guide de EMR mise à jour d'Amazon.
- La EMR version 5.32.0 d'Amazon est compatible avec Amazon EMR on. EKS Pour en savoir plus sur la prise en main EMR de EKS On, consultez [What is Amazon EMR on EKS](#).
- La EMR version 5.32.0 d'Amazon prend en charge Amazon EMR Studio (version préliminaire). Pour en savoir plus sur la prise en main de EMR Studio, consultez [Amazon EMR Studio \(version préliminaire\)](#).
- Politiques gérées délimitées : pour s'aligner sur les AWS meilleures pratiques, Amazon EMR a introduit des politiques gérées par défaut EMR définies dans la version v2 en remplacement des politiques qui seront déconseillées. Consultez les [politiques EMR gérées par Amazon](#).

Problèmes connus

- Pour les clusters de sous-réseaux privés Amazon EMR 6.3.0 et 6.2.0, vous ne pouvez pas accéder à l'interface utilisateur Web de Ganglia. Vous recevrez un message d'erreur « accès refusé (403) ». D'autres sites WebUIs, tels que Spark, Hue JupyterHub, Zeppelin, Livy et Tez, fonctionnent

normalement. L'accès à l'interface utilisateur Web de Ganglia sur les clusters de sous-réseaux publics fonctionne également normalement. Pour résoudre ce problème, redémarrez le service `httpd` sur le nœud primaire avec `sudo systemctl restart httpd`. Ce problème est résolu dans Amazon EMR 6.4.0.

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` et `emr-6.2.0` sont basées sur les anciennes versions de Linux of Amazon 2 (), qui ont un paramètre `ulimit` inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre `ulimit` EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre `ulimit` inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre `ulimit` lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'`ulimit` du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un `ulimit` à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section `Service`.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer `InstanceController`

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

Important

EMR les clusters qui exécutent Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) utilisent le comportement par défaut d'Amazon Linux et ne téléchargent ni n'installent automatiquement les mises à jour importantes et critiques du noyau nécessitant un redémarrage. Ce comportement est identique à celui des autres EC2 instances Amazon qui exécutent Amazon Linux par défaut AMI. Si de nouvelles mises à jour logicielles Amazon Linux nécessitant un redémarrage (telles que le noyau et les CUDA mises à jour) sont disponibles après la EMR sortie d'une version d'Amazon, les instances de EMR cluster qui exécutent la version par défaut AMI ne téléchargent ni n'installent automatiquement ces mises à jour. NVIDIA Pour obtenir les mises à jour du noyau, vous pouvez [personnaliser votre Amazon EMR AMI](#) pour qu'il [utilise la dernière version d'Amazon Linux AMI](#).

- La prise en charge par console pour créer une configuration de sécurité spécifiant l'option d'intégration de AWS Ranger n'est actuellement pas prise en charge dans la GovCloud région. La configuration de sécurité peut être effectuée à l'aide du CLI. Consultez la section [Création EMR de la configuration de sécurité](#) dans le guide EMR de gestion Amazon.
- Lorsque AtRestEncryption HDFS le chiffrement est activé sur un cluster qui utilise Amazon EMR 5.31.0 ou 5.32.0, les requêtes Hive entraînent l'exception d'exécution suivante.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
  attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
  java.lang.RuntimeException: Hive Runtime Error while closing
  operators: java.io.IOException: java.util.ServiceConfigurationError:
  org.apache.hadoop.security.token.TokenIdentifier: Provider
  org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :

 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.32.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon

EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,16,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.13.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-notebook-env</code>	1.1.0	Environnement Conda pour le bloc-notes EMR qui inclut la passerelle Jupyter Entreprise
<code>emr-s3-dist-cp</code>	2.17.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.6.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2,45,0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.11.2	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.11.2	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.1-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.1-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.1-amzn-0	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.10.1-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.1-amzn-0	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-ftpfs-server	2.10.1-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.1-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.1-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.1-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.1-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.7-amzn-3	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.7-amzn-3	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.7-amzn-3	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.7-amzn-3	Client de ligne de commande Hive.
hive-hbase	2.3.7-amzn-3	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.7-amzn-3	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.7-amzn-3	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.6.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.6.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.6.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.8.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.7.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5,5,68	Mon serveur SQL de base de données.
nvidia-cuda	10,1243	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.240.1-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.240.1-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.240.1-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.7-amzn-0	Clients de ligne de commande Spark.
spark-history-server	2.4.7-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.7-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.4.7-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	2.3.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.32.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Les actions de reconfiguration se produisent lorsque vous spécifiez une configuration pour les groupes d'instances d'un cluster en cours d'exécution. Amazon lance des actions de reconfiguration EMR uniquement pour les classifications que vous modifiez. Pour plus d'informations, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Classifications emr-5.32.0

Classifications	Description	Actions de reconfiguration
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.	Restarts the ResourceManager service.
container-executor	Modifiez les valeurs dans le fichier container-executor.cfg YARN de Hadoop.	Not available.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.	Not available.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Modifiez les paramètres liés au docker.	Not available.
emrfs-site	Modifiez EMRFS les paramètres.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classifications	Description	Actions de reconfiguration
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Modifiez les paramètres flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Modifiez les paramètres log4j.properties Flink.	Restarts Flink history server.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.	Not available.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.	Restarts Flink history server.

Classifications	Description	Actions de reconfiguration
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts PhoenixQueryserver, HiveServer2, Hive MetaStore, and MapReduce-HistoryServer.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Hadoop Httpfs, and MapReduce-HistoryServer.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop	Not available.
hadoop-ssl-client	Modifier la configuration du client ssl hadoop	Not available.

Classifications	Description	Actions de reconfiguration
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Changez les valeurs dans HBase l'environnement.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.	Not available.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.	Should not be reconfigured.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classifications	Description	Actions de reconfiguration
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.	Not available.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.	Not available.
hive-env	Modifiez les valeurs dans l'environnement Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.

Classifications	Description	Actions de reconfiguration
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.	Not available.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.	Not available.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2	Not available.
hue-ini	Modifiez les valeurs dans le fichier ini de Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.	Restarts Hadoop Httpfs service.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.	Not available.

Classifications	Description	Actions de reconfiguration
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.	Not available.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.	Not available.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.	Not available.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.	Not available.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.	Not available.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.	Not available.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.	Restarts Livy Server.
livy-env	Modifiez les valeurs dans l'environnement Livy.	Restarts Livy Server.
livy-log4j	Modifiez les paramètres Livy log4j.properties.	Restarts Livy Server.

Classifications	Description	Actions de reconfiguration
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.	Restarts Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.	Restarts Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.	Not available.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.	Not available.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.	Not available.
pig-env	Modifiez les valeurs dans l'environnement Pig.	Not available.

Classifications	Description	Actions de reconfiguration
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.	Restarts Oozie.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.	Not available.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.	Restarts Presto-Server.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.	Restarts Presto-Server.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.	Not available.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.	Restarts Presto-Server.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.	Not available.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.	Not available.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.	Not available.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.	Restarts Presto-Server.

Classifications	Description	Actions de reconfiguration
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.	Not available.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.	Not available.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.	Not available.
presto-connector-memory	Modifiez les valeurs dans le fichier <code>memory.properties</code> de Presto.	Not available.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.	Not available.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.	Not available.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.	Not available.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.	Not available.
presto-connector-redis	Modifiez les valeurs dans le fichier <code>redis.properties</code> de Presto.	Not available.

Classifications	Description	Actions de reconfiguration
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.	Not available.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.	Not available.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.	Not available.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site .xml de RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS	Not available.
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.	Not available.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.	Restarts EMR record server.

Classifications	Description	Actions de reconfiguration
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier <code>erver.properties</code> .	Restarts EMR record server.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier <code>log4j.properties</code> .	Restarts EMR record server.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Modifiez les valeurs dans le fichier <code>spark-defaults.conf</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Modifiez les valeurs dans l'environnement Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Modifiez les valeurs dans le fichier <code>hive-site.xml</code> de Spark	Not available.
spark-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Modifiez les valeurs dans le fichier <code>metrics.properties</code> de Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.	Not available.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier <code>oraoop-site.xml</code> de OraOop Sqoop.	Not available.

Classifications	Description	Actions de reconfiguration
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.	Not available.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.	Restarts Oozie.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.	Restarts Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.	Restarts Zookeeper server.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.	Restarts Zookeeper server.

Amazon EMR version 5.31.1

Versions d'application 5.31.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
Kit AWS SDK pour Java	1,1,852	1,1,852	1,1,759	1,1,759
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.11.0	1.10.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
HCatalog	2.3.7	2.3.7	2.3.6	2.3.6
Hadoop	2.10.0	2.10.0	2.8.5	2.8.5
Hive	2.3.7	2.3.7	2.3.6	2.3.6
Hudi	0.6.0-amzn-0	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating
Hue	4.7.1	4.7.1	4.6.0	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.6.0	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,238,3	0,238,3	0.232	0.232
Spark	2.4.6	2.4.6	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.1.0	2.1.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.31.1

Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.

Modifications, améliorations et problèmes résolus

- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état du nœud, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket

Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.

- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite maximale de fichiers ouverts sur les anciennes versions AL2 d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPC politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages.](#)

Problèmes connus

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode.](#)

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.31.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.4.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4,15,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.13.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.15.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.6.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2.43.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.11.0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.11.0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.0-amzn-0.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.0-amzn-0.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.0-amzn-0.1	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.10.0-amzn-0.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.0-amzn-0.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-ftpfs-server	2.10.0-amzn-0.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.10.0-amzn-0.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.0-amzn-0.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.0-amzn-0.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.0-amzn-0.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.0-amzn-0.1	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.7-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.7-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.7-amzn-1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.7-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.7-amzn-1	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.7-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.7-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.6.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-spark	0.6.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.6.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.7.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.6.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.3.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.238.3-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.238.3-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.238.3-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.6-amzn-0.1	Clients de ligne de commande Spark.
spark-history-server	2.4.6-amzn-0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.6-amzn-0.1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.4.6-amzn-0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	2.1.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.31.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.31.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.

Classifications	Description
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog <code>proto-hive-site</code> fichier <code>.xml</code> .
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat <code>log4j2.properties</code> de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier <code>webhcat-site.xml</code> de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier <code>beeline-log4j2.properties</code> de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.

Classifications	Description
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier <code>oozie-log4j.properties</code> d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier <code>oozie-site.xml</code> d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.

Classifications	Description
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.

Classifications	Description
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.31.0

Versions d'application 5.31.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
Kit AWS SDK pour Java	1,1,852	1,1,759	1,1,759	1,1,759
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.11.0	1.10.0	1.10.0	1.10.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.13
HCatalog	2.3.7	2.3.6	2.3.6	2.3.6
Hadoop	2.10.0	2.8.5	2.8.5	2.8.5
Hive	2.3.7	2.3.6	2.3.6	2.3.6
Hudi	0.6.0-amzn-0	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
Hue	4.7.1	4.6.0	4.6.0	4.6.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.1.0
Livy	0.7.0	0.7.0	0.7.0	0.7.0
MXNet	1.6.0	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.2.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0,238,3	0.232	0.232	0.232
Spark	2.4.6	2.4.5	2.4.5	2.4.5
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	2.1.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.31.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.31.0 d'Amazon. Les modifications ont été apportées à la version 5.30.1.

Date de parution initiale : 9 octobre 2020

Dernière mise à jour : 15 octobre 2020

Mises à niveau

- Mise à niveau du connecteur Amazon Glue vers la version 1.13.0
- Mise à niveau d'Amazon SageMaker Spark SDK vers la version 1.4.0
- Mise à niveau du connecteur Amazon Kinesis vers la version 3.5.9
- Mise à niveau AWS SDK for Java vers la version 1.11.852
- Mise à niveau de Bigtop-tomcat vers la version 8.5.56
- EMRFS mis à niveau vers la version 2.43.0
- EMR MetricsAndEventsApiGateway Client mis à niveau vers la version 1.4.0
- Mise à niveau de EMR S3 Dist CP vers la version 2.15.0
- EMRS3 Select mis à niveau vers la version 1.6.0
- Mise à niveau de Flink vers la version 1.11.0
- Mise à niveau de Hadoop vers la version 2.10.0
- Mise à niveau de Hive vers la version 2.3.7
- Mise à niveau de Hudi vers la version 0.6.0
- Mise à niveau de Hue vers la version 4.7.1
- Mise à niveau JupyterHub vers la version 1.1.0
- Mise à niveau de Mxnet vers la version 1.6.0
- Mise à niveau d'OpenCV vers la version 4.3.0
- Mise à niveau de Presto vers la version 0.238.3
- Mise à niveau TensorFlow vers la version 2.1.0

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état du nœud, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite maximale de fichiers ouverts sur les anciennes versions AL2 d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Les [statistiques des colonnes Hive](#) sont prises en charge pour les EMR versions 5.31.0 et ultérieures d'Amazon.
- Versions de composants mises à niveau.

- EMRFSSupport du S3EC V2 dans Amazon EMR 5.31.0. Dans les SDK versions 1.11.837 et ultérieures de S3 Java, le client de chiffrement version 2 (S3EC V2) a été introduit avec diverses améliorations de sécurité. Pour plus d'informations, consultez les ressources suivantes :
 - Article de blog S3 : [Updates to the Amazon S3 encryption client](#).
 - AWS SDK for Java Guide du développeur : [Migrez les clients de chiffrement et de déchiffrement vers la version V2](#).
 - EMRGuide de gestion : [chiffrement côté client Amazon S3](#).

Le client de chiffrement V1 est toujours disponible dans le SDK pour des raisons de rétrocompatibilité.

Nouvelles fonctionnalités

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMRLes versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux ofAmazon 2 (), qui ont un paramètre ulimit inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMRLes versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre ulimit EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latestAmazon AMI Le paramètre ulimit inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre ulimit lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'ulimit du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un ulimit à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section Service.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Avec Amazon EMR 5.31.0, vous pouvez lancer un cluster qui s'intègre à Lake Formation. Cette intégration fournit un filtrage des données précis au niveau des colonnes pour les bases de données et les tables du Glue AWS Data Catalog. Il permet également l'authentification unique fédérée à EMR Notebooks ou Apache Zeppelin à partir d'un système d'identité d'entreprise. Pour plus d'informations, consultez la section [Intégration EMR d'Amazon AWS Lake Formation](#) dans le guide EMR de gestion Amazon.

Amazon EMR with Lake Formation est actuellement disponible dans 16 AWS régions : USA Est (Ohio et Virginie du Nord), USA Ouest (Californie du Nord et Oregon), Asie-Pacifique (Mumbai, Séoul, Singapour, Sydney et Tokyo), Canada (Centre), Europe (Francfort, Irlande, Londres, Paris et Stockholm), Amérique du Sud (São Paulo).

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Lorsque AtRestEncryption HDFS le chiffrement est activé sur un cluster qui utilise Amazon EMR 5.31.0 ou 5.32.0, les requêtes Hive entraînent l'exception d'exécution suivante.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :

- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.31.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.4.0	Amazon SageMaker Spark SDK
emr-ddb	4,15,0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.13.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.15.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.6.0	EMRConnecteur S3Select
emrfs	2.43.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.11.0	Applications et scripts client de la ligne de commande Apache Flink.
flink-jobmanager-config	1.11.0	Gestion des ressources sur les EMR nœuds pour Apache Flink JobManager.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.10.0-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.10.0-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.10.0-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.10.0-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.10.0-amzn-0	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.10.0-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.10.0-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.10.0-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.10.0-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.10.0-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.10.0-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.7-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.7-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.7-amzn-1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.7-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.7-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.7-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.7-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.6.0-amzn-0	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.

Composant	Version	Description
hudi-spark	0.6.0-amzn-0	Bibliothèque de solution groupée pour exécuter Spark avec Hudi.
hudi-presto	0.6.0-amzn-0	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.7.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.6.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	4.3.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.238.3-amzn-0	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.238.3-amzn-0	Service pour exécuter les éléments d'une requête.
presto-client	0.238.3-amzn-0	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques

Composant	Version	Description
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.6-amzn-0	Clients de ligne de commande Spark.
spark-history-server	2.4.6-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.6-amzn-0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.6-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	2.1.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.31.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.31.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.

Classifications	Description
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-PROTO-HIVE-SITE	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2

Classifications	Description
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.

Classifications	Description
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.

Classifications	Description
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.

Classifications	Description
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.

Classifications	Description
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.30.2

Versions d'application 5.30.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.30.2	emr-5.30.1	emr-5.30.0	mr-5.29.0
Kit AWS SDK pour Java	1,1,759	1,1,759	1,1,759	1,11,682
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.10.0	1.9.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.13	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.2-incubating	0.5.2-incubating	0.5.0-incubating
Hue	4.6.0	4.6.0	4.6.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.1.0	1.0.0
Livy	0.7.0	0.7.0	0.7.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1

	emr-5.30.2	emr-5.30.1	emr-5.30.0	mr-5.29.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.2.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.232	0.232	0.227
Spark	2.4.5	2.4.5	2.4.5	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.30.2

Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.

Modifications, améliorations et problèmes résolus

- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état du nœud, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas

en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR

- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPCE politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages.](#)

Problèmes connus

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement

les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :

- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+, -`. Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.30.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK

Composant	Version	Description
emr-ddb	4.14.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.13.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.14.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.5.0	EMRConnecteur S3Select
emrfs	2.40.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.10.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-6.1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-6.1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-6.1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-6.1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-6.1	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	2.8.5-amzn-6.1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-6.1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-6.1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.8.5-amzn-6.1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-6.1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-6.1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.6-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.6-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.6-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.6-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.6-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.2-incubating	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.5.2-incubating	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.6.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.5.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.232	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.232	Service pour exécuter les éléments d'une requête.
presto-client	0.232	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.5-amzn-0.1	Clients de ligne de commande Spark.
spark-history-server	2.4.5-amzn-0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	2.4.5-amzn-0.1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.4.5-amzn-0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.30.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.30.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>yarn-session</code> fichier <code>.properties</code> de Flink <code>log4j</code> .
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop

Classifications	Description
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.

Classifications	Description
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.

Classifications	Description
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.

Classifications	Description
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.

Classifications	Description
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.

Classifications	Description
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.

Classifications	Description
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.30.1

Versions d'application 5.30.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.30.1	emr-5.30.0	mr-5.29.0	emr-5.28.1
Kit AWS SDK pour Java	1,1,759	1,1,759	1,11,682	1,1,659
Python	2,7, 3,7	2,7, 3,7	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.10.0	1.9.1	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.13	1.4.10	1.4.10

	emr-5.30.1	emr-5.30.0	mr-5.29.0	emr-5.28.1
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.2-incubating	0.5.0-incubating	0.5.0-incubating
Hue	4.6.0	4.6.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.1.0	1.0.0	1.0.0
Livy	0.7.0	0.7.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.2.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.232	0.227	0.227
Spark	2.4.5	2.4.5	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2

	emr-5.30.1	emr-5.30.0	mr-5.29.0	emr-5.28.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.30.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.30.1 d'Amazon. Les modifications ont été apportées à la version 5.30.0.

Date de parution initiale : 30 juin 2020

Dernière mise à jour : 24 août 2020

Modifications, améliorations et problèmes résolus

- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Correction d'un problème à cause duquel le processus du contrôleur d'instance générait un nombre infini de processus.
- Correction d'un problème où Hue ne parvenait pas à exécuter une requête Hive, affichant un message « la base de données est verrouillée » et empêchant l'exécution des requêtes.
- Correction d'un problème lié à Spark qui permettait à un plus grand nombre de tâches de s'exécuter simultanément sur le EMR cluster.
- Correction d'un problème de bloc-notes Jupyter provoquant une erreur « trop de fichiers ouverts » dans le serveur Jupyter.
- Correction d'un problème lié aux temps de démarrage des clusters.

Nouvelles fonctionnalités

- Les interfaces d'application persistantes de Tez UI et de YARN Timeline Server sont disponibles avec EMR les versions 6.x, 5.30.1 et EMR ultérieures d'Amazon. L'accès par lien en un clic à l'historique permanent des candidatures vous permet d'accéder rapidement à l'historique des tâches sans configurer de proxy Web via une SSH connexion. Les journaux des clusters actifs et résiliés sont disponibles pendant 30 jours après la fin de l'application. Pour plus d'informations, consultez la section [Afficher les interfaces utilisateur persistantes des applications](#) dans le Amazon EMR Management Guide.
- EMR L'exécution APIs de blocs-notes est disponible pour exécuter EMR des blocs-notes via un script ou une ligne de commande. La possibilité de démarrer, d'arrêter, de répertorier et de décrire les exécutions de EMR blocs-notes sans la AWS console vous permet de contrôler un EMR bloc-notes par programme. À l'aide d'une cellule de bloc-notes paramétrée, vous pouvez transmettre différentes valeurs de paramètres à un bloc-notes sans avoir à créer une copie du bloc-notes pour chaque nouvel ensemble de valeurs de paramètres. Voir [EMRAPIActions](#). Pour un exemple de code, voir [Exemples de commandes pour exécuter des EMR blocs-notes par programmation](#).

Problèmes connus

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux ofAmazon 2 (), qui ont un paramètre ulimit inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre ulimit EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latestAmazon AMI Le paramètre ulimit inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre ulimit lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'ulimit du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un ulimit à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section `Service`.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer `InstanceController`

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
```

```
sudo systemctl daemon-reload
```

- EMRCarnets

La fonctionnalité qui permet d'installer des noyaux et des bibliothèques Python supplémentaires sur le nœud principal du cluster est désactivée par défaut dans la EMR version 5.30.1. Pour plus d'informations sur cette fonctionnalité, consultez [Installation de noyaux et de bibliothèques Python sur un nœud primaire de cluster](#).

Pour activer cette fonctionnalité, procédez comme suit :

1. Assurez-vous que la politique d'autorisation associée au rôle de service pour les EMR blocs-notes autorise les actions suivantes :

```
elasticmapreduce:ListSteps
```

Pour plus d'informations, voir [Rôle de service pour les EMR ordinateurs portables](#).

2. Utilisez le AWS CLI pour exécuter une étape sur le cluster qui configure les EMR blocs-notes, comme indiqué dans l'exemple suivant. Remplacez *us-east-1* avec la région dans laquelle réside votre cluster. Pour plus d'informations sur l'ajout d'étapes, consultez la rubrique [Ajout d'étapes à un cluster à l'aide de la AWS CLI](#).

```
aws emr add-steps --cluster-id MyClusterID --steps
  Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
```

- Mise à l'échelle gérée

Les opérations de mise à l'échelle gérées sur des clusters 5.30.0 et 5.30.1 sans Presto installé peuvent provoquer des défaillances d'applications ou empêcher le maintien d'un groupe d'instances ou d'une flotte d'instances uniforme dans l'état ARRESTED, en particulier lorsqu'une opération de réduction est rapidement suivie d'une opération d'augmentation.

Pour contourner le problème, choisissez Presto comme application à installer lorsque vous créez un cluster avec les EMR versions 5.30.0 et 5.30.1 d'Amazon, même si votre travail ne nécessite pas Presto.

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.

- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.30.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.13.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.14.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.5.0	EMRConnecteur S3Select
emrfs	2.40.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.10.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-6	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».

Composant	Version	Description
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-6	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.8.5-amzn-6	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-6	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-6	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.8.5-amzn-6	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.6-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.6-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	2.3.6-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.6-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.6-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.2-incubating	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.5.2-incubating	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.6.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.5.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.232	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.232	Service pour exécuter les éléments d'une requête.
presto-client	0.232	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.5-amzn-0	Clients de ligne de commande Spark.
spark-history-server	2.4.5-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.5-amzn-0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.4.5-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.30.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.30.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.

Classifications	Description
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog <code>proto-hive-site</code> fichier <code>.xml</code> .
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat <code>log4j2.properties</code> de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier <code>webhcat-site.xml</code> de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier <code>beeline-log4j2.properties</code> de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.

Classifications	Description
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier <code>oozie-log4j.properties</code> d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier <code>oozie-site.xml</code> d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.

Classifications	Description
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.

Classifications	Description
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.30.0

Versions d'application 5.30.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.30.0	mr-5.29.0	emr-5.28.1	emr-5.28.0
Kit AWS SDK pour Java	1,1,759	1,11,682	1,1,659	1,1,659
Python	2,7, 3,7	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.10.0	1.9.1	1.9.0	1.9.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.13	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.6
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.6
Hudi	0.5.2-incubating	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating

	emr-5.30.0	mr-5.29.0	emr-5.28.1	emr-5.28.0
Hue	4.6.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.1.0	1.0.0	1.0.0	1.0.0
Livy	0.7.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.5.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.2.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.3
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.232	0.227	0.227	0.227
Spark	2.4.5	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.2
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.30.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.30.0 d'Amazon. Les modifications ont été apportées à la version 5.29.0.

Date de parution initiale : 13 mai 2020

Date de la dernière mise à jour : 25 juin 2020

Mises à niveau

- Mise à niveau AWS SDK for Java vers la version 1.11.759
- Mise à niveau d'Amazon SageMaker Spark SDK vers la version 1.3.0
- Serveur d'EMR enregistrement mis à niveau vers la version 1.6.0
- Mise à niveau de Flink vers la version 1.10.0
- Mise à niveau de Ganglia vers la version 3.7.2
- Mise à niveau HBase vers la version 1.4.13
- Mise à niveau de Hudi vers la version 0.5.2 incubating
- Mise à niveau de Hue vers la version 4.6.0
- Mise à niveau JupyterHub vers la version 1.1.0
- Mise à niveau de Livy vers la version 0.7.0-incubating
- Mise à niveau d'Oozie vers la version 5.2.0
- Mise à niveau de Presto vers la version 0.232
- Mise à niveau de Spark vers la version 2.4.5
- Connecteurs et pilotes améliorés : Amazon Glue Connector 1.12.0 ; Amazon Kinesis Connector 3.5.0 ; DynamoDB Connector 4.14.0 EMR

Nouvelles fonctionnalités

- EMR Blocs-notes — Lorsqu'ils sont utilisés avec des EMR clusters créés à l'aide de la version 5.30.0, les noyaux des EMR blocs-notes s'exécutent sur le cluster. Cela améliore les performances des blocs-notes et vous permet d'installer et de personnaliser les noyaux. Vous pouvez également installer des bibliothèques Python sur le nœud primaire du cluster. Pour plus d'informations, consultez la section [Installation et utilisation des noyaux et des bibliothèques](#) dans le Guide EMR de gestion.

- Dimensionnement géré : avec Amazon EMR version 5.30.0 et versions ultérieures, vous pouvez activer le dimensionnement EMR géré pour augmenter ou diminuer automatiquement le nombre d'instances ou d'unités dans votre cluster en fonction de la charge de travail. Amazon évalue EMR en permanence les indicateurs des clusters afin de prendre des décisions de dimensionnement qui optimisent les coûts et la rapidité de vos clusters. Pour plus d'informations, consultez la section [Scaling Cluster Resources](#) dans le guide EMR de gestion Amazon.
- Chiffrer les fichiers journaux stockés dans Amazon S3 : avec Amazon EMR version 5.30.0 et versions ultérieures, vous pouvez chiffrer les fichiers journaux stockés dans Amazon S3 à l'aide d'une AWS KMS clé gérée par le client. Pour plus d'informations, consultez la section [Chiffrer les fichiers journaux stockés dans Amazon S3](#) dans le guide EMR de gestion Amazon.
- Support d'Amazon Linux 2 : dans les EMR versions 5.30.0 et ultérieures, EMR uses Amazon système d'exploitation Linux 2. La nouvelle personnalisation AMIs (Amazon Machine Image) doit être basée sur the Amazon Linux 2 AMI. Pour plus d'informations, consultez la section [Utilisation d'une personnalisation AMI](#).
- Presto Graceful Auto Scale : les EMR clusters utilisant la version 5.30.0 peuvent être définis avec un délai de mise à l'échelle automatique qui donne aux tâches Presto le temps de terminer leur exécution avant que leur nœud ne soit mis hors service. Pour plus d'informations, consultez [Utilisation du dimensionnement automatique de Presto avec désaffectation gracieuse](#).
- Création d'une instance de flotte avec une nouvelle option de stratégie d'allocation — Une nouvelle option de stratégie d'allocation est disponible dans les EMR versions 5.12.1 et ultérieures. Cela permet un approvisionnement plus rapide des clusters, une allocation plus précise et moins d'interruptions d'instances Spot. Des mises à jour des rôles de EMR service autres que ceux par défaut sont requises. Consultez [Configuration de parcs d'instances](#).
- Commandes `sudo systemctl stop` et `sudo systemctl start` : dans les EMR versions 5.30.0 et ultérieures, quel système d'exploitation use Amazon Linux 2 utilise et commande pour redémarrer les services. `EMR sudo systemctl stop sudo systemctl start` Pour plus d'informations, consultez [Comment redémarrer un service sur Amazon EMR ?](#).

Modifications, améliorations et problèmes résolus

- EMR la version 5.30.0 n'installe pas Ganglia par défaut. Lorsque vous créez un cluster, vous pouvez sélectionner expressément l'installation de Ganglia.
- Optimisation des performances Spark.
- Optimisation des performances Presto.
- Python 3 est la EMR version par défaut d'Amazon 5.30.0 et versions ultérieures.

- Le groupe de sécurité géré par défaut pour l'accès au service dans les sous-réseaux privés a été mis à jour avec de nouvelles règles. Si vous utilisez un groupe de sécurité personnalisé pour accéder au service, vous devez inclure les mêmes règles que le groupe de sécurité géré par défaut. Pour plus d'informations, consultez [Amazon EMR -Managed Security Group for Service Access \(sous-réseaux privés\)](#). Si vous utilisez un rôle de service personnalisé pour AmazonEMR, vous devez accorder l'autorisation `ec2:describeSecurityGroups` EMR afin de valider si les groupes de sécurité ont été correctement créés. Si vous utilisez le `EMR_DefaultRole`, cette autorisation est déjà incluse dans la stratégie gérée par défaut.

Problèmes connus

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` et `emr-6.2.0` sont basées sur les anciennes versions de Linux ofAmazon 2 (), qui ont un paramètre `ulimit` inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions `5.30.1`, `5.30.2`, `5.31.1`, `5.32.1`, `6.0.1`, `6.1.1`, `6.2.1`, `5.33.0`, `6.3.0` et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre `ulimit` EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latestAmazon AMI Le paramètre `ulimit` inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre `ulimit` lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'`ulimit` du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un `ulimit` à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section Service.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Mise à l'échelle gérée

Les opérations de mise à l'échelle gérées sur des clusters 5.30.0 et 5.30.1 sans Presto installé peuvent provoquer des défaillances d'applications ou empêcher le maintien d'un groupe d'instances ou d'une flotte d'instances uniforme dans l'état ARRESTED, en particulier lorsqu'une opération de réduction est rapidement suivie d'une opération d'augmentation.

Pour contourner le problème, choisissez Presto comme application à installer lorsque vous créez un cluster avec les EMR versions 5.30.0 et 5.30.1 d'Amazon, même si votre travail ne nécessite pas Presto.

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Le moteur de base de données par défaut pour Hue 4.6.0 est SQLite, ce qui pose des problèmes lorsque vous essayez d'utiliser Hue avec une base de données externe. Pour résoudre ce problème, définissez `engine` de votre configuration de classification `hue.ini` sur `mysql`. Ce problème a été résolu dans la EMR version 5.30.1 d'Amazon.

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Versions des composants 5.30.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.13.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.5.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-notebook-env	1.0.0	Environnement Conda pour bloc-notes emr
emr-s3-dist-cp	2.14.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.5.0	EMRConnecteur S3Select
emrfs	2.40.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.10.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-6	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-6	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	2.8.5-amzn-6	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-6	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.8.5-amzn-6	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-6	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.13	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.13	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.13	HBaseclient en ligne de commande.
hbase-rest-server	1.4.13	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.13	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hcatalog-client	2.3.6-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.6-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.6-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.6-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.6-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.2-incubating	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.5.2-incubating	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.

Composant	Version	Description
hue-server	4.6.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.1.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.7.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.5.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mariadb-server	5.5.64	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.2.0	Client de ligne de commande Oozie.
oozie-server	5.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur

Composant	Version	Description
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.232	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.232	Service pour exécuter les éléments d'une requête.
presto-client	0.232	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.3	Projet R pour les calculs statistiques
ranger-kms-server	1.2.0	Système de gestion des clés Apache Ranger
spark-client	2.4.5-amzn-0	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.4.5-amzn-0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.5-amzn-0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.5-amzn-0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.30.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.30.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.

Classifications	Description
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
hudi-env	Modifiez les valeurs dans l'environnement Hudi.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Classifications	Description
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.29.0

Versions d'application 5.29.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	mr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
Kit AWS SDK pour Java	1,11,682	1,1,659	1,1,659	1,1,615
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12

	mr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.1	1.9.0	1.9.0	1.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.6	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.6	2.3.5
Hudi	0.5.0-incubating	0.5.0-incubating	0.5.0-incubating	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.5.1	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.3	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	mr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
Presto	0.227	0.227	0.227	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.2	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.29.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.29.0 d'Amazon. Les modifications ont été apportées à la version 5.28.1.

Date de parution initiale : 17 janvier 2020

Mises à niveau

- Mise à niveau AWS SDK for Java vers la version 1.11.682
- Mise à niveau de Hive vers la version 2.3.6
- Mise à niveau de Flink vers la version 1.9.1
- Mise à niveau d'EmrFS vers la version 2.38.0
- Connecteur EMR DynamoDB mis à niveau vers la version 4.13.0

Modifications, améliorations et problèmes résolus

- Spark
 - Optimisation des performances Spark.

- EMRFS
 - Mise à jour du guide de gestion avec les paramètres par défaut du fichier `emrfs-site.xml` pour un affichage cohérent.

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'`hadooputilisateur` du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier `keytab` se trouve dans `/etc/hadoop.keytab` et le `principal` se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.29.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.13.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.12.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.13.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.4.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2.38.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.9.1	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-5	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-5	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httfs-server	2.8.5-amzn-5	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-5	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.10	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.10	Service pour desservir une ou plusieurs HBase régions.

Composant	Version	Description
hbase-client	1.4.10	HBaseclient en ligne de commande.
hbase-rest-server	1.4.10	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.10	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.6-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.6-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.6-amzn-1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.6-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.6-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server2	2.3.6-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.0-incubating	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.5.0-incubating	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.0.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.5.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.

Composant	Version	Description
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.227	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.227	Service pour exécuter les éléments d'une requête.
presto-client	0.227	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.4	Clients de ligne de commande Spark.
spark-history-server	2.4.4	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.4	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.4	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.29.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.29.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.

Classifications	Description
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-PROTO-HIVE-SITE	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2

Classifications	Description
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.

Classifications	Description
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.28.1

Versions d'application 5.28.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
Kit AWS SDK pour Java	1,1,659	1,1,659	1,1,615	1,1,615
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.0	1.9.0	1.8.1	1.8.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.6	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.6	2.3.5	2.3.5
Hudi	0.5.0-incubating	0.5.0-incubating	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	1.0.0
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.5.1	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.3	4.14.2	4.14.2

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.227	0.227	0.224	0.224
Spark	2.4.4	2.4.4	2.4.4	2.4.4
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.14.0
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.2	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.28.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.28.1 d'Amazon. Les modifications ont été apportées à la version 5.28.0.

Date de parution initiale : 10 janvier 2020

Modifications, améliorations et problèmes résolus

- Spark
 - Correction des problèmes de compatibilité Spark.
- CloudWatch Métriques
 - Correction de la publication d'Amazon CloudWatch Metrics sur un EMR cluster comportant plusieurs nœuds principaux.
- Message de journal désactivé
 - Message de faux journal désactivé, « ...en utilisant l'ancienne version (<4.5.8) du client HTTP Apache ».

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.28.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version

d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.12.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.11.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.13.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.3.0	EMRConnecteur S3Select
emrfs	2.37.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.9.0	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-5	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-5	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	2.8.5-amzn-5	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.8.5-amzn-5	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.10	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.10	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.10	HBaseclient en ligne de commande.
hbase-rest-server	1.4.10	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.4.10	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.6-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.6-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.6-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.6-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.6-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.6-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.0-incubating	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.

Composant	Version	Description
hudi-presto	0.5.0-incubating	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.0.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.5.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.227	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.227	Service pour exécuter les éléments d'une requête.
presto-client	0.227	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.4	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.4.4	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.4	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.4	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.28.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.28.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.

Classifications	Description
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Classifications	Description
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.28.0

Versions d'application 5.28.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#), et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
Kit AWS SDK pour Java	1,1,659	1,1,615	1,1,615	1,1595
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.9.0	1.8.1	1.8.1	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.10
HCatalog	2.3.6	2.3.5	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.6	2.3.5	2.3.5	2.3.5
Hudi	0.5.0-incubating	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	1.0.0	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.5.1	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.3	4.14.2	4.14.2	4.14.2
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
Presto	0.227	0.224	0.224	0.220
Spark	2.4.4	2.4.4	2.4.4	2.4.3
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.14.0	1.13.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.2	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.28.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.28.0 d'Amazon. Les modifications ont été apportées à la version 5.27.0.

Date de parution initiale : 12 novembre 2019

Mises à niveau

- Mise à niveau de Flink vers la version 1.9.0
- Mise à niveau de Hive vers la version 2.3.6
- Mise à niveau MXNet vers la version 1.5.1
- Mise à niveau de Phoenix vers la version 4.14.3
- Mise à niveau de Presto vers la version 0.227
- Mise à niveau de Zeppelin vers la version 0.8.2

Nouvelles fonctionnalités

- Amazon EMR peut désormais installer [Apache Hudi](#) lorsque vous créez un cluster. Pour plus d'informations, consultez [Hudi](#).
- (25 novembre 2019) Vous pouvez maintenant choisir d'exécuter plusieurs étapes en parallèle pour améliorer l'utilisation du cluster et faire des économies. Vous pouvez également annuler à la fois les étapes en attente et celles en cours d'exécution. Pour plus d'informations, voir [Travailler avec des étapes à l'aide de la console AWS CLI AND](#).
- (3 décembre 2019) Vous pouvez désormais créer et exécuter des EMR clusters sur AWS Outposts. AWS Outposts permet AWS des services, une infrastructure et des modèles d'exploitation natifs dans les installations sur site. Dans AWS Outposts les environnements, vous pouvez utiliser les mêmes AWS APIs outils et infrastructures que ceux que vous utilisez dans le AWS cloud. Pour plus d'informations, consultez la section [EMRclusters sur AWS Outposts](#).
- (11 mars 2020) À partir de la EMR version 5.28.0 d'Amazon, vous pouvez créer et exécuter des EMR clusters Amazon sur un sous-réseau de zones AWS locales en tant qu'extension logique d'une AWS région qui prend en charge les zones locales. Une zone locale permet aux EMR fonctionnalités d'Amazon et à un sous-ensemble de AWS services, tels que les services de calcul et de stockage, d'être situés plus près des utilisateurs, offrant ainsi un accès à très faible latence aux applications exécutées localement. Pour obtenir la liste des zones locales disponibles, consultez [AWS Local Zones](#). Pour plus d'informations sur l'accès aux zones AWS locales disponibles, voir [Régions, zones de disponibilité et zones locales](#).

Les zones Locales ne prennent actuellement pas en charge les EMR ordinateurs portables Amazon et ne prennent pas en charge les connexions directes à Amazon EMR via l'interface VPC endpoint (AWS PrivateLink).

Modifications, améliorations et problèmes résolus

- Prise en charge d'applications étendues pour des clusters à haute disponibilité
 - Pour plus d'informations, consultez la section [Applications prises en charge dans un EMR cluster comportant plusieurs nœuds principaux](#) dans le guide EMR de gestion Amazon.
- Spark
 - Optimisation des performances
- Hive
 - Optimisation des performances

- Presto
 - Optimisation des performances

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.28.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.11.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.13.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.3.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2.37.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.9.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-5	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-5	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httfs-server	2.8.5-amzn-5	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-5	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.10	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.10	Service pour desservir une ou plusieurs HBase régions.

Composant	Version	Description
hbase-client	1.4.10	HBaseclient en ligne de commande.
hbase-rest-server	1.4.10	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.10	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.6-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.6-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.6-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.6-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.6-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.6-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server2	2.3.6-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hudi	0.5.0-incubating	Infrastructure de traitement incrémentiel pour implémenter un pipeline à faible latence et à efficacité élevée.
hudi-presto	0.5.0-incubating	Bibliothèque de solution groupée pour exécuter Presto avec Hudi.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.0.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.5.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.

Composant	Version	Description
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,3- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,3- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.227	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.227	Service pour exécuter les éléments d'une requête.
presto-client	0.227	Client de ligne de commande Presto installé sur les maîtres de secours d'un cluster haute disponibilité où le serveur Presto n'est pas démarré.

Composant	Version	Description
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.4	Clients de ligne de commande Spark.
spark-history-server	2.4.4	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.4	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.4	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.2	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.28.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.28.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.

Classifications	Description
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2

Classifications	Description
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.pr operties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-en v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.prop erties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole .properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra .properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.prop erties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.prope rties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.pro perties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile .properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.pr operties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.p roperties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.pro perties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.

Classifications	Description
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.27.1

Versions d'application 5.27.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Kit AWS SDK pour Java	1,1,615	1,1,615	1,1595	1,1,566
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.1	1.8.1	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.10	1.4.9
HCatalog	2.3.5	2.3.5	2.3.5	2.3.5
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.5	2.3.5
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	1.0.0	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.2	4.14.1

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.224	0.224	0.220	0.220
Spark	2.4.4	2.4.4	2.4.3	2.4.3
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.14.0	1.13.1	1.13.1
Tez	0.9.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.14

Notes de mise à jour 5.27.1

Il s'agit d'une version de correctif. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

État de prise en charge du service de métadonnées d'instance (IMDS) V2 : les composants Amazon EMR 5.23.1, 5.27.1 et 5.32 ou versions ultérieures sont utilisés pour tous les appels. IMDSv2 IMDS Pour les IMDS appels dans le code de votre application, vous pouvez utiliser les deux IMDSv1 etIMDSv2, ou les configurer pour qu'IMDSils ne soient utilisés que IMDSv2 pour renforcer la sécurité. Pour les autres EMR versions 5.x, la désactivation IMDSv1 entraîne l'échec du démarrage du cluster.

Versions des composants 5.27.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark SDK
emr-ddb	4.12.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.11.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.13.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.3.0	EMRConnecteur S3Select
emrfs	2.36.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.8.1	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème

Composant	Version	Description
		ème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httpfs-server	2.8.5-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.8.5-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.10	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.10	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.10	HBaseclient en ligne de commande.
hbase-rest-server	1.4.10	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.4.10	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.5-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.5-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.5-amzn-1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.5-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.5-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.5-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.5-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
jupyterhub	1.0.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.4.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,2- -1,4 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,14,2- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.224	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.224	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.4	Clients de ligne de commande Spark.
spark-history-server	2.4.4	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.4	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.4	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.27.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.27.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.

Classifications	Description
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog <code>proto-hive-site</code> fichier <code>.xml</code> .
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat <code>log4j2.properties</code> de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier <code>webhcat-site.xml</code> de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier <code>beeline-log4j2.properties</code> de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier <code>oozie-log4j.properties</code> d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier <code>oozie-site.xml</code> d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.

Classifications	Description
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.

Classifications	Description
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.27.0

Versions d'application 5.27.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
Kit AWS SDK pour Java	1,1,615	1,1595	1,1,566	1,1,546
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.1	1.8.0	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.10	1.4.9	1.4.9
HCatalog	2.3.5	2.3.5	2.3.5	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.5	2.3.4
Hudi	-	-	-	-

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	1.0.0	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.2	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.224	0.220	0.220	0.219
Spark	2.4.4	2.4.3	2.4.3	2.4.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.14.0	1.13.1	1.13.1	1.12.0
Tez	0.9.2	0.9.2	0.9.2	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.14	3.4.13

Notes de mise à jour 5.27.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.27.0 d'Amazon. Les modifications ont été apportées à la version 5.26.0.

Date de parution initiale : 23 septembre 2019

Mises à niveau

- AWS SDK for Java 1,1,615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- TensorFlow 1.14.0
- Connecteurs et pilotes :
 - Connecteur DynamoDB 4.12.0

Nouvelles fonctionnalités

- (24 octobre 2019) Les nouvelles fonctionnalités suivantes des EMR ordinateurs portables sont disponibles avec toutes les EMR versions d'Amazon.
 - Vous pouvez désormais associer des référentiels Git à des EMR blocs-notes pour stocker vos blocs-notes dans un environnement à version contrôlée. Vous pouvez partager du code avec vos pairs et réutiliser les blocs-notes Jupyter existants via des référentiels Git distants. Pour plus d'informations, consultez [Associer des référentiels Git à Amazon EMR Notebooks](#) dans le guide de EMRgestion Amazon.
 - L'[utilitaire nbdime](#) est désormais disponible dans les ordinateurs portables pour simplifier la comparaison et la fusion EMR des blocs-notes.
 - EMRles ordinateurs portables sont désormais compatibles JupyterLab. JupyterLab est un environnement de développement interactif basé sur le Web entièrement compatible avec les ordinateurs portables Jupyter. Vous pouvez désormais choisir d'ouvrir votre bloc-notes dans l'éditeur de bloc-notes Jupyter JupyterLab ou dans l'éditeur de bloc-notes Jupyter.
- (30 octobre 2019) Avec EMR les versions 5.25.0 et ultérieures d'Amazon, vous pouvez vous connecter à l'interface utilisateur du serveur d'historique Spark depuis la page de résumé du cluster ou depuis l'onglet Historique des applications de la console. Au lieu de configurer un proxy Web par le biais d'une SSH connexion, vous pouvez accéder rapidement à l'interface utilisateur du

serveur d'historique Spark pour consulter les métriques de l'application et accéder aux fichiers journaux pertinents pour les clusters actifs et interrompus. Pour plus d'informations, consultez la section [Accès hors cluster aux interfaces utilisateur persistantes des applications](#) dans le Amazon EMR Management Guide.

Modifications, améliorations et problèmes résolus

- EMRCluster Amazon avec plusieurs nœuds principaux
 - Vous pouvez installer et exécuter Flink sur un EMR cluster Amazon comportant plusieurs nœuds principaux. Pour plus d'informations, consultez [Applications et fonctionnalités prises en charge](#).
 - Vous pouvez configurer le chiffrement HDFS transparent sur un EMR cluster Amazon comportant plusieurs nœuds principaux. Pour plus d'informations, consultez la section [Chiffrement HDFS transparent sur les EMR clusters comportant plusieurs nœuds principaux](#).
 - Vous pouvez désormais modifier la configuration des applications exécutées sur un EMR cluster Amazon comportant plusieurs nœuds principaux. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).
- Connecteur Amazon EMR -DynamoDB
 - Le connecteur Amazon EMR -DynamoDB prend désormais en charge les types de données DynamoDB suivants : boolean, list, map, item, null. Pour plus d'informations, consultez [Configuration d'une table Hive pour l'exécution de commandes Hive](#).

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.27.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark SDK

Composant	Version	Description
emr-ddb	4.12.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.11.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.13.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.3.0	EMRConnecteur S3Select
emrfs	2.36.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.8.1	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.8.5-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.10	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.10	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.10	HBaseclient en ligne de commande.
hbase-rest-server	1.4.10	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.10	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.5-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.5-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.5-amzn-1	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.5-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.5-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.5-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.5-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	1.0.0	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.4.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,2- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,2- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.224	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.224	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.4	Clients de ligne de commande Spark.
spark-history-server	2.4.4	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.4	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.4	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.14.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.

Composant	Version	Description
zeppelin-server	0.8.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.27.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.27.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .

Classifications	Description
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.

Classifications	Description
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.

Classifications	Description
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.

Classifications	Description
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.

Classifications	Description
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.

Classifications	Description
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
ranger-kms-dbks-site	Modifiez les valeurs dans le fichier dbks-site.xml de RangerKMS.
ranger-kms-site	Modifiez les valeurs dans le fichier ranger-kms-site.xml de RangerKMS.
ranger-kms-env	Modifiez les valeurs dans l'KMSenvironnement Ranger.
ranger-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Ranger. KMS
ranger-kms-db-ca	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.

Classifications	Description
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.26.0

Versions d'application 5.26.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
Kit AWS SDK pour Java	1,1595	1,1,566	1,1,546	1,1,546
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.8.0	1.8.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.10	1.4.9	1.4.9	1.4.9
HCatalog	2.3.5	2.3.5	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.5	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.4.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.6	0.9.6
Livy	0.6.0	0.6.0	0.6.0	0.6.0
MXNet	1.4.0	1.4.0	1.4.0	1.4.0

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.2	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.220	0.220	0.219	0.219
Spark	2.4.3	2.4.3	2.4.2	2.4.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.13.1	1.13.1	1.12.0	1.12.0
Tez	0.9.2	0.9.2	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.14	3.4.13	3.4.13

Notes de mise à jour 5.26.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.26.0 d'Amazon. Les modifications ont été apportées à la version 5.25.0.

Date de parution initiale : 8 août 2019

Dernière mise à jour : 19 août 2019

Mises à niveau

- AWS SDK for Java 1,1595
- HBase1.4.10
- Phoenix 4.14.2

- Connecteurs et pilotes :
 - Connecteur DynamoDB 4.11.0
 - Connecteur MariaDB 2.4.2
 - Pilote Amazon Redshift JDBC 1.2.32.1056

Nouvelles fonctionnalités

- (Bêta) Avec Amazon EMR 5.26.0, vous pouvez lancer un cluster qui s'intègre à Lake Formation. Cette intégration fournit un accès détaillé au niveau des colonnes aux bases de données et aux tables du Glue AWS Data Catalog. Il permet également l'authentification unique fédérée à EMR Notebooks ou Apache Zeppelin à partir d'un système d'identité d'entreprise. Pour plus d'informations, consultez [Intégrer Amazon EMR à AWS Lake Formation \(version bêta\)](#).
- (19 août 2019) L'accès public EMR bloqué par Amazon est désormais disponible dans toutes les EMR versions d'Amazon qui prennent en charge les groupes de sécurité. Bloquer l'accès public est un paramètre applicable à l'ensemble du compte et appliqué à chaque AWS région. Le blocage de l'accès public empêche le lancement d'un cluster lorsqu'un groupe de sécurité associé au cluster possède une règle autorisant le trafic entrant depuis IPv4 0.0.0.0/0 ou IPv6 : :/0 (accès public) sur un port, sauf si un port est spécifié comme exception. Le port 22 est une exception par défaut. Pour plus d'informations, consultez la section [Utilisation d'Amazon EMR Block Public Access](#) dans le guide EMR de gestion Amazon.

Modifications, améliorations et problèmes résolus

- Blocs-notes EMR
 - À partir de la version EMR 5.26.0, EMR Notebooks prend en charge les bibliothèques Python adaptées aux ordinateurs portables en plus des bibliothèques Python par défaut. Vous pouvez installer des bibliothèques limitées au bloc-notes depuis l'éditeur de bloc-notes sans avoir à recréer un cluster ou à rattacher un bloc-notes à un cluster. Les bibliothèques limitées aux blocs-notes sont créées dans un environnement virtuel Python, de sorte qu'elles ne s'appliquent qu'à la session de bloc-notes en cours. Cela vous permet d'isoler les dépendances du bloc-notes. Pour plus d'informations, consultez la section [Using Notebook Scoped Libraries](#) dans le Amazon EMR Management Guide.
- EMRFS
 - Vous pouvez activer une fonctionnalité ETag de vérification (version bêta) en réglant `fs.s3.consistent.metadata.etag.validation.enabled` sur `true`. Grâce à cette

fonctionnalité, EMRFS utilise Amazon S3 ETags pour vérifier que les objets en cours de lecture sont de la dernière version disponible. Cette fonctionnalité est utile dans les cas read-after-update d'utilisation dans lesquels des fichiers sur Amazon S3 sont remplacés tout en conservant le même nom. Cette fonctionnalité de ETag vérification ne fonctionne pas actuellement avec S3 Select. Pour plus d'informations, consultez [Configuration de la vue cohérente](#).

- Spark
 - Les optimisations suivantes sont désormais activées par défaut : élagage dynamique des partitions, amélioration de l'inférence des statistiques DISTINCT du SQL plan pour les requêtes JOIN suivies INTERSECT, aplatissement des DISTINCT sous-requêtes scalaires, réorganisation optimisée des jointures et jointure par filtre Bloom. Pour plus d'informations, consultez [Optimisation des performances de Spark](#).
 - Génération de code améliorée pour l'ensemble de l'étape pour Sort Merge Join.
 - Amélioration de la réutilisation des fragments de requête et des sous-requêtes.
 - Améliorations apportées à la préallocation des exécuteurs au démarrage de Spark.
 - Les jointures par filtre bloom ne sont plus appliquées lorsque le côté le plus petit de la jointure inclut un indice de diffusion.
- Tez
 - Un problème avec Tez a été résolu. L'interface utilisateur de Tez fonctionne désormais sur un EMR cluster Amazon avec plusieurs nœuds principaux.

Problèmes connus

- Les fonctionnalités améliorées de génération de code pour l'ensemble de l'étape pour Sort Merge Join peuvent augmenter la sollicitation de la mémoire lorsqu'elles sont activées. Cette optimisation améliore les performances, mais peut entraîner de nouvelles tentatives de tâches ou des échecs si `spark.yarn.executor.memoryOverheadFactor` n'est pas réglé pour fournir suffisamment de mémoire. Pour désactiver cette fonctionnalité, définissez `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` sur `false`.
- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la

période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.26.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark SDK
emr-ddb	4.11.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.10.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.12.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.3.0	EMRConnecteur S3Select
emrfs	2.35.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.8.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.8.5-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.10	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.10	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.10	HBaseclient en ligne de commande.
hbase-rest-server	1.4.10	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.10	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.5-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.5-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.5-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.5-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.5-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.5-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.5-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.6	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.4.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,2- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,2- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.220	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.220	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.3	Clients de ligne de commande Spark.
spark-history-server	2.4.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.3	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.13.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.

Composant	Version	Description
zeppelin-server	0.8.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.26.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.26.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .

Classifications	Description
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.

Classifications	Description
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.

Classifications	Description
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.

Classifications	Description
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.

Classifications	Description
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.

Classifications	Description
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.

Classifications	Description
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.25.0

Versions d'application 5.25.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
Kit AWS SDK pour Java	1,1,566	1,1,546	1,1,546	1,1,519
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.8.0	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
HCatalog	2.3.5	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.5	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.4.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.6	0.9.4
Livy	0.6.0	0.6.0	0.6.0	0,5,0
MXNet	1.4.0	1.4.0	1.4.0	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.220	0.219	0.219	0.215
Spark	2.4.3	2.4.2	2.4.2	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.13.1	1.12.0	1.12.0	1.12.0
Tez	0.9.2	0.9.1	0.9.1	0.9.1

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.14	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.25.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.25.0 d'Amazon. Les modifications ont été apportées à la version 5.24.1.

Date de parution initiale : 17 juillet 2019

Dernière mise à jour : 30 octobre 2019

Amazon EMR 5.25.0

Mises à niveau

- AWS SDK for Java 1,1,566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3
- TensorFlow 1.13.1
- Tez 0.9.2
- Zookeeper 3.4.14

Nouvelles fonctionnalités

- (30 octobre 2019) À partir de la EMR version 5.25.0 d'Amazon, vous pouvez vous connecter à l'interface utilisateur du serveur d'historique Spark depuis la page de résumé du cluster ou depuis l'onglet Historique des applications de la console. Au lieu de configurer un proxy Web par le biais d'une SSH connexion, vous pouvez accéder rapidement à l'interface utilisateur du serveur d'historique Spark pour consulter les métriques de l'application et accéder aux fichiers journaux pertinents pour les clusters actifs et interrompus. Pour plus d'informations, consultez la section

[Accès hors cluster aux interfaces utilisateur persistantes des applications](#) dans le Amazon EMR Management Guide.

Modifications, améliorations et problèmes résolus

- Spark
 - Amélioration des performances de certaines jointures en utilisant des filtres Bloom pour préfiltrer les entrées. L'optimisation est désactivée par défaut et peut être activée en définissant le paramètre de configuration Spark `spark.sql.bloomFilterJoin.enabled` sur `true`.
 - Amélioration des performances du regroupement par colonnes de type chaîne.
 - Amélioration de la mémoire par défaut de l'exécuteur Spark et de la configuration des cœurs des types d'instances R4 pour les clusters non HBase installés.
 - Résolution d'un problème antérieur lié à la fonctionnalité d'élimination dynamique des partitions, où la table éliminée devait se trouver du côté gauche de la jointure.
 - Amélioré DISTINCT avant INTERSECT l'optimisation pour s'appliquer à d'autres cas impliquant des alias.
 - Amélioration de l'inférence des statistiques du SQL plan pour les requêtes JOIN suivies par les DISTINCT requêtes. Cette amélioration est désactivée par défaut et peut être activée en définissant le paramètre de configuration Spark `spark.sql.statsImprovements.enabled` sur `true`. Cette optimisation est requise par la fonctionnalité Distinct avant Intersect et sera automatiquement activée lorsque `spark.sql.optimizer.distinctBeforeIntersect.enabled` est définie sur `true`.
 - Ordre de jointure optimisé en fonction de la taille de la table et des filtres. Cette optimisation est désactivée par défaut et peut être activée en définissant le paramètre de configuration Spark `spark.sql.optimizer.sizeBasedJoinReorder.enabled` sur `true`.

Pour plus d'informations, consultez [Optimisation des performances de Spark](#).

- EMRFS
 - Le EMRFS paramètre est désormais désactivé par défaut. `fs.s3.buckets.create.enabled` Lors des tests, nous avons constaté que la désactivation de ce paramètre améliore les performances et empêche la création involontaire de compartiments S3. Si votre application repose sur cette fonctionnalité, vous pouvez l'activer en définissant la propriété `fs.s3.buckets.create.enabled` sur `true` dans la classification de configuration `emrfs-site`. Pour plus d'informations, consultez [Fourniture d'une configuration lors de la création d'un cluster](#).

- Améliorations du chiffrement du disque local et du chiffrement S3 dans les configurations de sécurité (5 août 2019)
 - Les paramètres de chiffrement Amazon S3 ont été séparés des paramètres de chiffrement du disque local dans la configuration de la sécurité.
 - Ajout d'une option pour activer EBS le chiffrement avec les versions 5.24.0 et ultérieures. La sélection de cette option chiffre le volume du périphérique racine en plus des volumes de stockage. Les versions précédentes nécessitaient l'utilisation d'une AMI méthode personnalisée pour chiffrer le volume du périphérique racine.
 - Pour plus d'informations, consultez la section [Options de chiffrement](#) dans le guide EMR de gestion Amazon.

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.25.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.10.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.9.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.11.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.3.0	EMRConnecteur S3Select
emrfs	2.34.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.8.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».

Composant	Version	Description
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.8.5-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.9	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.9	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.9	HBaseclient en ligne de commande.
hbase-rest-server	1.4.9	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.9	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.5-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.5-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.5-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	2.3.5-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.5-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.5-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.5-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.6	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.4.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.

Composant	Version	Description
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14.1- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14.1- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.220	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.220	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques

Composant	Version	Description
spark-client	2.4.3	Clients de ligne de commande Spark.
spark-history-server	2.4.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.3	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.13.1	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.2	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.1	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.14	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.14	ZooKeeper client en ligne de commande.

Classifications des configurations 5.25.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.25.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.

Classifications	Description
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2

Classifications	Description
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
recordserver-env	Modifiez les valeurs dans l'EMR RecordServer environnement.
recordserver-conf	Modifiez les valeurs dans le EMR RecordServer fichier erver.properties.
recordserver-log4j	Modifiez les valeurs dans le EMR RecordServer fichier log4j.properties.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark

Classifications	Description
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.24.1

Versions d'application 5.24.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
Kit AWS SDK pour Java	1,1,546	1,1,546	1,1,519	1,1,519
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.8.0	1.7.1	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.4.0	4.4.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.6	0.9.4	0.9.4
Livy	0.6.0	0.6.0	0,5,0	0,5,0
MXNet	1.4.0	1.4.0	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.219	0.219	0.215	0.215
Spark	2.4.2	2.4.2	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.24.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.24.1 d'Amazon. Les modifications ont été apportées à la version 5.24.0.

Date de parution initiale : 26 juin 2019

Modifications, améliorations et problèmes résolus

- Mise à jour de la version par défaut d'Amazon Linux AMI pour Amazon EMR afin d'inclure d'importantes mises à jour de sécurité du noyau Linux, notamment le problème de TCP SACK déni de service ([AWS-2019-005](#)).

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.24.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion-amzn-EmrVersion.EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK

Composant	Version	Description
emr-ddb	4.9.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.8.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.11.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.3.0	EMRConnecteur S3Select
emrfs	2.33.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.8.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.8.5-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.9	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.9	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.9	HBaseclient en ligne de commande.
hbase-rest-server	1.4.9	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.9	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.4-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.4-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.4-amzn-2	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.4-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.4-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.6	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.4.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14.1- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14.1- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.219	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.219	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.2	Clients de ligne de commande Spark.
spark-history-server	2.4.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.

Composant	Version	Description
zeppelin-server	0.8.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.24.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.24.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .

Classifications	Description
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.

Classifications	Description
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.

Classifications	Description
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.

Classifications	Description
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.

Classifications	Description
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.

Classifications	Description
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.24.0

Versions d'application 5.24.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
Kit AWS SDK pour Java	1,1,546	1,1,519	1,1,519	1,11,510
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.8.0	1.7.1	1.7.1	1.7.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.9
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
Hue	4.4.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.6	0.9.4	0.9.4	0.9.4
Livy	0.6.0	0,5,0	0,5,0	0,5,0
MXNet	1.4.0	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.1.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.219	0.215	0.215	0.215
Spark	2.4.2	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.1
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.24.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.24.0 d'Amazon. Les modifications ont été apportées à la version 5.23.0.

Date de parution initiale : 11 juin 2019

Dernière mise à jour : 5 août 2019

Mises à niveau

- Flink 1.8.0
- Hue 4.4.0
- JupyterHub 0,9,6
- Livy 0.6.0
- MxNet 1.4.0
- Presto 0.219
- Spark 2.4.2
- AWS SDK for Java 1,1,546
- Connecteurs et pilotes :
 - Connecteur DynamoDB 4.9.0
 - Connecteur MariaDB 2.4.1
 - Pilote Amazon Redshift JDBC 1.2.27.1051

Modifications, améliorations et problèmes résolus

- Spark
 - Ajout d'une optimisation pour éliminer dynamiquement les partitions. L'optimisation est désactivée par défaut. Pour l'activer, définissez le paramètre de configuration Spark `spark.sql.dynamicPartitionPruning.enabled` sur `true`.
 - Amélioration des performances des requêtes INTERSECT. L'optimisation est désactivée par défaut. Pour l'activer, définissez le paramètre de configuration Spark `spark.sql.optimizer.distinctBeforeIntersect.enabled` sur `true`.
 - Ajout d'une optimisation pour aplatir les sous-requêtes scalaires avec des agrégats qui utilisent la même relation. L'optimisation est désactivée par défaut. Pour l'activer, définissez le paramètre de configuration Spark

```
spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled sur true.
```

- Génération de code améliorée pour l'ensemble de l'étape.

Pour plus d'informations, consultez [Optimisation des performances de Spark](#).

- Améliorations du chiffrement du disque local et du chiffrement S3 dans les configurations de sécurité (5 août 2019)
 - Les paramètres de chiffrement Amazon S3 ont été séparés des paramètres de chiffrement du disque local dans la configuration de la sécurité.
 - Ajout d'une option pour activer EBS le chiffrement. La sélection de cette option chiffre le volume du périphérique racine en plus des volumes de stockage. Les versions précédentes nécessitaient l'utilisation d'une AMI méthode personnalisée pour chiffrer le volume du périphérique racine.
 - Pour plus d'informations, consultez la section [Options de chiffrement](#) dans le guide EMR de gestion Amazon.

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.24.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.9.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-goodies	2.8.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.11.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.3.0	EMRConnecteur S3Select
emrfs	2.33.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.8.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.8.5-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.8.5-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.9	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.9	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.9	HBaseclient en ligne de commande.
hbase-rest-server	1.4.9	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.9	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.4-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.4-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.

Composant	Version	Description
hcatalog-webhcat-server	2.3.4-amzn-2	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.4-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.4-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.4.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.6	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.6.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.

Composant	Version	Description
mxnet	1.4.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14.1- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14.1- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.219	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.219	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.2	Clients de ligne de commande Spark.
spark-history-server	2.4.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.1	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.24.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.24.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.

Classifications	Description
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-PROTO-HIVE-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2

Classifications	Description
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.23.1

Versions d'application 5.23.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
Kit AWS SDK pour Java	1,1,519	1,1,519	1,11,510	1,11,479
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.1	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.9	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.1.0	5.0.0
Phoenix	4.14.1	4.14.1	4.14.1	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.1	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.23.1

Il s'agit d'une version de correctif. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

État de prise en charge du service de métadonnées d'instance (IMDS) V2 : les composants Amazon EMR 5.23.1, 5.27.1 et 5.32 ou versions ultérieures sont utilisés pour tous les appels. IMDSv2 IMDS Pour les IMDS appels dans le code de votre application, vous pouvez utiliser les deux IMDSv1 et IMDSv2, ou les configurer pour qu'IMDS ils ne soient utilisés que IMDSv2 pour renforcer la sécurité. Pour les autres EMR versions 5.x, la désactivation IMDSv1 entraîne l'échec du démarrage du cluster.

Versions des composants 5.23.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.8.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.7.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-s3-dist-cp	2.11.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.2.0	EMRConnecteur S3Select
emrfs	2.32.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.7.1	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.8.5-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-https-server	2.8.5-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	1.4.9	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.9	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.9	HBaseclient en ligne de commande.
hbase-rest-server	1.4.9	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.9	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.4-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.4-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.4-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-1	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.4-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.4-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.3.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.3.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda

Composant	Version	Description
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14.1- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14.1- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.215	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.215	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.0	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.4.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.4.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.23.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.23.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.

Classifications	Description
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.

Classifications	Description
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.23.0

Versions d'application 5.23.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)

- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
Kit AWS SDK pour Java	1,1,519	1,11,510	1,11,479	1,11,479
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.1	1.7.1	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.9	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.1.0	5.0.0	5.0.0
Phoenix	4.14.1	4.14.1	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.1	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.23.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.23.0 d'Amazon. Les modifications ont été apportées à la version 5.22.0.

Date de parution initiale : 1 avril 2019

Dernière mise à jour : 30 avril 2019

Mises à niveau

- AWS SDK for Java 1,1,519

Nouvelles fonctionnalités

- (30 avril 2019) Avec Amazon EMR 5.23.0 et versions ultérieures, vous pouvez lancer un cluster avec trois nœuds principaux pour prendre en charge la haute disponibilité d'applications telles que YARN Resource Manager, Spark HDFS NameNode, Hive et Ganglia. Le nœud primaire n'est plus un point de défaillance potentiel grâce à cette fonctionnalité. Si l'un des nœuds principaux tombe en panne, Amazon EMR bascule automatiquement vers un nœud principal de secours et remplace le nœud principal défaillant par un nouveau nœud présentant les mêmes actions de configuration et de démarrage. Pour plus d'informations, consultez [Planification et configuration des nœuds primaires](#).

Problèmes connus

- Interface utilisateur Tez (corrigée dans la EMR version 5.26.0 d'Amazon)

L'interface utilisateur de Tez ne fonctionne pas sur un EMR cluster comportant plusieurs nœuds principaux.

- Hue (corrigé dans la EMR version 5.24.0 d'Amazon)
 - Hue exécuté sur Amazon EMR ne prend pas en charge Solr. À compter de la EMR version 5.20.0 d'Amazon, un problème de configuration incorrecte entraîne l'activation de Solr et un message d'erreur inoffensif similaire au suivant s'affiche :

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

Pour empêcher l'affichage du message d'erreur de Solr, procédez comme suit :

1. Connectez-vous à la ligne de commande du nœud principal à l'aide deSSH.
2. Utilisez un éditeur de texte pour ouvrir le fichier hue.ini. Par exemple :

```
sudo vim /etc/hue/conf/hue.ini
```

3. Recherchez le terme `appblacklist` et modifiez la ligne comme suit :

```
appblacklist = search
```

4. Enregistrez vos modifications et redémarrez Hue comme indiqué dans l'exemple suivant :

```
sudo stop hue; sudo start hue
```

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'`hadooputilisateur` du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier `keytab` se trouve dans `/etc/hadoop.keytab` et le `principal` se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.23.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.8.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.7.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.2.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2.32.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.7.1	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFSservice de gestion du journal du système de fichiers Hadoop sur les clusters HA.
hadoop-httfs-server	2.8.5-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.9	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.9	Service pour desservir une ou plusieurs HBase régions.

Composant	Version	Description
hbase-client	1.4.9	HBaseclient en ligne de commande.
hbase-rest-server	1.4.9	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.9	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.4-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.4-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.4-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server2	2.3.4-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.3.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.3.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14.1- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14.1- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.215	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.215	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.0	Clients de ligne de commande Spark.
spark-history-server	2.4.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.0	Moteur d'exécution en mémoire pour YARN

Composant	Version	Description
spark-yarn-slave	2.4.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.23.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.23.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.

Classifications	Description
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.

Classifications	Description
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.

Classifications	Description
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.

Classifications	Description
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.

Classifications	Description
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.

Classifications	Description
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.22.0

Versions d'application 5.22.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
Kit AWS SDK pour Java	1,11,510	1,11,479	1,11,479	1,11,479
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.1	1.7.0	1.7.0	1.7.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.9	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.1.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.1	4.14.0	4.14.0	4.14.0

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.215
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.1	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.22.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.22.0 d'Amazon. Les modifications ont été apportées à la version 5.21.0.

Important

Depuis la EMR version 5.22.0 d'Amazon, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. EMR Les versions antérieures d'Amazon utilisent AWS la version 2 de Signature dans certains cas, sauf si les notes de publication indiquent que la version 4 de Signature est utilisée exclusivement. Pour plus d'informations, consultez les [sections Authentification des demandes \(AWS Signature version 4\)](#) et [Authentification des demandes \(AWS Signature version 2\)](#) dans le manuel Amazon Simple Storage Service Developer Guide.

Date de parution initiale : 20 mars 2019

Mises à niveau

- Flink 1.7.1
- HBase1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Connecteurs et pilotes :
 - Connecteur DynamoDB 4.8.0
 - Connecteur MariaDB 2.2.6
 - Pilote Amazon Redshift JDBC 1.2.20.1043

Nouvelles fonctionnalités

- Modification de la EBS configuration par défaut pour les types d'EC2instances avec stockage EBS uniquement. Lorsque vous créez un cluster à l'aide d'Amazon EMR version 5.22.0 ou ultérieure, la quantité de EBS stockage par défaut augmente en fonction de la taille de l'instance. En outre, nous avons réparti le stockage accru sur plusieurs volumes, ce qui a permis d'améliorer les performances. Si vous souhaitez utiliser une autre configuration de stockage d'EBSinstance, vous pouvez la spécifier lorsque vous créez un EMR cluster ou que vous ajoutez des nœuds à un cluster existant. Pour plus d'informations sur la quantité de stockage et le nombre de volumes alloués par défaut pour chaque type d'instance, consultez la section [EBSStockage par défaut pour les instances](#) dans le guide EMR de gestion Amazon.

Modifications, améliorations et problèmes résolus

- Spark
 - Ajout d'une nouvelle propriété de configuration pour Spark `onYARN,spark.yarn.executor.memoryOverheadFactor`. La valeur de cette propriété est un facteur d'échelle qui définit la valeur de la surcharge mémoire à un pourcentage de la mémoire de l'exécuteur, avec un minimum de 384 Mo. Si la surcharge mémoire est définie explicitement à l'aide `spark.yarn.executor.memoryOverhead`, cette propriété n'a aucun effet. La valeur par défaut est `0.1875`, ce qui représente 18,75 %. Cette valeur par défaut pour Amazon EMR laisse plus d'espace dans les YARN conteneurs pour la surcharge de mémoire de l'exécuteur que la valeur par défaut de 10 % définie en interne par Spark. La EMR valeur par

défaut d'Amazon de 18,75 % a montré empiriquement moins de défaillances liées à la mémoire dans les benchmarks -DS. TPC

- Rétroporté [SPARK-26316](#) pour améliorer les performances.
- Dans les EMR versions 5.19.0, 5.20.0 et 5.21.0 d'Amazon, les étiquettes des YARN nœuds sont stockées dans un répertoire. HDFS Dans certaines situations, cela entraîne des retards dans le démarrage des nœuds principaux, ce qui provoque le dépassement du délai du cluster et l'échec du lancement. À partir d'Amazon EMR 5.22.0, ce problème est résolu. YARN les étiquettes des nœuds sont stockées sur le disque local de chaque nœud du cluster, ce qui permet d'éviter toute dépendance vis-à-vis de HDFS.

Problèmes connus

- Hue (corrigé dans la EMR version 5.24.0 d'Amazon)
 - Hue exécuté sur Amazon EMR ne prend pas en charge Solr. À compter de la EMR version 5.20.0 d'Amazon, un problème de configuration incorrecte entraîne l'activation de Solr et un message d'erreur inoffensif similaire au suivant s'affiche :

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

Pour empêcher l'affichage du message d'erreur de Solr, procédez comme suit :

1. Connectez-vous à la ligne de commande du nœud principal à l'aide de SSH.
2. Utilisez un éditeur de texte pour ouvrir le fichier `hue.ini`. Par exemple :

```
sudo vim /etc/hue/conf/hue.ini
```

3. Recherchez le terme `appblacklist` et modifiez la ligne comme suit :

```
appblacklist = search
```

4. Enregistrez vos modifications et redémarrez Hue comme indiqué dans l'exemple suivant :

```
sudo stop hue; sudo start hue
```

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters dotés de plusieurs nœuds principaux et d'une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois que le cluster est en cours d'exécution depuis un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.22.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version

d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.8.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.6.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.11.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.2.0	EMRConnecteur S3Select
emrfs	2.31.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.7.1	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.8.5-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.8.5-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.9	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.9	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.9	HBaseclient en ligne de commande.
hbase-rest-server	1.4.9	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.9	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hcatalog-client	2.3.4-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.4-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.4-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.4-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.3.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark

Composant	Version	Description
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.3.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.1.0	Client de ligne de commande Oozie.
oozie-server	5.1.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14.1- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14.1- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.215	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.215	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.0	Clients de ligne de commande Spark.
spark-history-server	2.4.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.

Composant	Version	Description
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.22.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.22.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.

Classifications	Description
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.

Classifications	Description
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.

Classifications	Description
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.

Classifications	Description
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.

Classifications	Description
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.

Classifications	Description
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.21.2

Versions d'application 5.21.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
Kit AWS SDK pour Java	1,11,479	1,11,479	1,11,479	1,11,461
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.7.0	1.7.0	1.6.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.8

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.215	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.21.2

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.21.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.7.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.5.1	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.11.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.2.0	EMRConnecteur S3Select
emrfs	2.30.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.7.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.8.5-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.8	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.8	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.8	HBaseclient en ligne de commande.
hbase-rest-server	1.4.8	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.8	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.4-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.

Composant	Version	Description
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.4-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.4-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.3.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	REST interface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.

Composant	Version	Description
mxnet	1.3.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.215	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.215	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.0	Clients de ligne de commande Spark.
spark-history-server	2.4.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.0	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.21.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.21.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.

Classifications	Description
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-PROTO-HIVE-SITE	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2

Classifications	Description
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.21.1

Versions d'application 5.21.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
Kit AWS SDK pour Java	1,11,479	1,11,479	1,11,461	1,11,461
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2,1,12	2,1,12	2,1,12	2,1,12
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.7.0	1.6.2	1.6.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.8
HCatalog	2.3.4	2.3.4	2.3.4	2.3.4
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.4
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.1	1.3.1	1.3.1	1.3.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.215	0.214	0.214
Spark	2.4.0	2.4.0	2.4.0	2.4.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.12.0
Tez	0.9.1	0.9.1	0.9.1	0.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.21.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.21.1 d'Amazon. Les modifications ont été apportées à la version 5.21.0.

Date de parution initiale : 18 juillet 2019

Modifications, améliorations et problèmes résolus

- Mise à jour de la version par défaut d'Amazon Linux AMI pour Amazon EMR afin d'inclure d'importantes mises à jour de sécurité du noyau Linux, notamment le problème de TCP SACK déni de service ([AWS-2019-005](#)).

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.21.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.2.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2.30.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.7.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.8.5-amzn-1	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.8	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.8	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.8	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.4.8	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.4.8	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	2.3.4-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.4-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.4-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	4.3.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.3.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur

Composant	Version	Description
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.215	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.215	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.0	Clients de ligne de commande Spark.
spark-history-server	2.4.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.21.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.21.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.

Classifications	Description
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.

Classifications	Description
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.

Classifications	Description
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.

Classifications	Description
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.

Classifications	Description
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.

Classifications	Description
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.21.0

Versions d'application 5.21.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
Kit AWS SDK pour Java	1,11,479	1,11,461	1,11,461	1,1,433
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,4

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
Scala	2,1,12	2,1,12	2,1,12	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.7.0	1.6.2	1.6.2	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.8	1.4.7
HCatalog	2.3.4	2.3.4	2.3.4	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.4	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.3.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.1	1.3.1	1.3.1	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.215	0.214	0.214	0.212
Spark	2.4.0	2.4.0	2.4.0	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.12.0	1.11.0
Tez	0.9.1	0.9.1	0.9.1	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.21.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.21.0 d'Amazon. Les modifications ont été apportées à la version 5.20.0.

Date de parution initiale : 18 février 2019

Dernière mise à jour : 3 avril 2019

Mises à niveau

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1,11,479

Nouvelles fonctionnalités

- (3 avril 2019) Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier des classifications de configuration supplémentaires pour

chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, utilisez la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).

Modifications, améliorations et problèmes résolus

- Zeppelin
 - Rétroporté [ZEPPELIN-3878](#).

Problèmes connus

- Hue (corrigé dans la EMR version 5.24.0 d'Amazon)
 - Hue exécuté sur Amazon EMR ne prend pas en charge Solr. À compter de la EMR version 5.20.0 d'Amazon, un problème de configuration incorrecte entraîne l'activation de Solr et un message d'erreur inoffensif similaire au suivant s'affiche :

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

Pour empêcher l'affichage du message d'erreur de Solr, procédez comme suit :

1. Connectez-vous à la ligne de commande du nœud principal à l'aide deSSH.
2. Utilisez un éditeur de texte pour ouvrir le fichier hue.ini. Par exemple :

```
sudo vim /etc/hue/conf/hue.ini
```

3. Recherchez le terme `appblacklist` et modifiez la ligne comme suit :

```
appblacklist = search
```

4. Enregistrez vos modifications et redémarrez Hue comme indiqué dans l'exemple suivant :

```
sudo stop hue; sudo start hue
```

- Tez

- Ce problème a été résolu dans Amazon EMR 5.22.0.

Lorsque vous vous connectez à l'interface utilisateur de Tez sur `http://MasterDNS:8080/tez-ui` via une SSH connexion au nœud principal du cluster, l'erreur « L'opération de l'adaptateur a échoué - Le serveur Timeline (ATS) est hors de portée. Soit il est en panne, soit il CORS est « non activé » s'affiche, soit les tâches affichent « N/A » de manière inattendue.

Cela est dû au fait que l'interface utilisateur de Tez envoie des demandes au serveur YARN Timeline en utilisant `localhost` plutôt que le nom d'hôte du nœud principal. Pour contourner le problème, un script est disponible pour être exécuté en tant qu'action ou étape d'amorçage. Le script met à jour le nom d'hôte dans le fichier Tez `configs.env`. Pour plus d'informations et pour connaître l'emplacement du script, consultez les [Instructions sur l'amorçage](#).

- Dans les EMR versions 5.19.0, 5.20.0 et 5.21.0 d'Amazon, les étiquettes des YARN nœuds sont stockées dans un répertoire. HDFS Dans certaines situations, cela entraîne des retards dans le démarrage des nœuds principaux, ce qui provoque le dépassement du délai du cluster et l'échec du lancement. À partir d'Amazon EMR 5.22.0, ce problème est résolu. YARN Les étiquettes des nœuds sont stockées sur le disque local de chaque nœud du cluster, ce qui permet d'éviter toute dépendance vis-à-vis de HDFS.
- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSH en tant qu'hadoop utilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.21.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-goodies	2.5.1	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.11.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.2.0	EMRConnecteur S3Select
emrfs	2.30.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.7.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.8.5-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.8.5-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.8	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.8	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.8	HBaseclient en ligne de commande.
hbase-rest-server	1.4.8	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.8	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.4-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	2.3.4-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.4-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.3.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.3.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.

Composant	Version	Description
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.215	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.215	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques

Composant	Version	Description
spark-client	2.4.0	Clients de ligne de commande Spark.
spark-history-server	2.4.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.0	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.21.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.21.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.

Classifications	Description
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-PROTO-HIVE-SITE	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2

Classifications	Description
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.20.1

Versions d'application 5.20.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
Kit AWS SDK pour Java	1,11,461	1,11,461	1,1,433	1,1,433
Python	2,7, 3,6	2,7, 3,6	2,7, 3,4	2,7, 3,4
Scala	2,1,12	2,1,12	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.2	1.6.2	1.6.1	1.6.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.8	1.4.7	1.4.7
HCatalog	2.3.4	2.3.4	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.5
Hive	2.3.4	2.3.4	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.3.0	4.2.0	4.2.0
Iceberg	-	-	-	-

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.9.4
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.1	1.3.1	1.3.0	1.3.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.214	0.212	0.212
Spark	2.4.0	2.4.0	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.12.0	1.11.0	1.11.0
Tez	0.9.1	0.9.1	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.13

Notes de mise à jour 5.20.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

⚠ Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.20.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-s3-dist-cp	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.2.0	EMRConnecteur S3Select
emrfs	2.29.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.6.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.8.5-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.8.5-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.8	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.4.8	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.8	HBaseclient en ligne de commande.
hbase-rest-server	1.4.8	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.8	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.4-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.4-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server2	2.3.4-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.3.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.3.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.214	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.214	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.0	Clients de ligne de commande Spark.
spark-history-server	2.4.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.4.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.20.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.20.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.

Classifications	Description
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.

Classifications	Description
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.

Classifications	Description
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.

Classifications	Description
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.

Classifications	Description
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.

Classifications	Description
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.20.0

Versions d'application 5.20.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
Kit AWS SDK pour Java	1,11,461	1,1,433	1,1,433	1,11,393
Python	2,7, 3,6	2,7, 3,4	2,7, 3,4	2,7, 3,4

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
Scala	2,1,12	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.2	1.6.1	1.6.1	1.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.8	1.4.7	1.4.7	1.4.7
HCatalog	2.3.4	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.5	2.8.4
Hive	2.3.4	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.3.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.9.4	0.8.1
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.1	1.3.0	1.3.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.214	0.212	0.212	0.210
Spark	2.4.0	2.3.2	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.12.0	1.11.0	1.11.0	1.9.0
Tez	0.9.1	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.13	3.4.12

Notes de mise à jour 5.20.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.20.0 d'Amazon. Les modifications ont été apportées à la version 5.19.0.

Date de parution initiale : 18 décembre 2018

Dernière mise à jour : 22 janvier 2019

Mises à niveau

- Flink 1.6.2
- HBase1.4.8
- Hive 2.3.4
- Hue 4.3.0
- MXNet1.3.1
- Presto 0.214

- Spark 2.4.0
- TensorFlow 1,12,0
- Tez 0.9.1
- AWS SDK for Java 1,11,461

Nouvelles fonctionnalités

- (22 janvier 2019) Kerberos dans Amazon EMR a été amélioré pour prendre en charge l'authentification des principaux depuis un outil externe. KDC Cela permet de centraliser la gestion principale, car plusieurs clusters peuvent partager un seul cluster externeKDC. En outre, l'entité externe KDC peut bénéficier d'une confiance entre domaines avec un domaine Active Directory. Cela permet à tous les clusters d'authentifier les mandataires à partir d'Active Directory. Pour plus d'informations, consultez [Utiliser l'authentification Kerberos](#) dans le guide de EMRgestion Amazon.

Modifications, améliorations et problèmes résolus

- Amazon Linux par défaut AMI pour Amazon EMR
 - Le package Python3 a été mis à jour de python 3.4 à 3.6.
- Le convertisseur EMRFS optimisé pour S3
 - Le validateur EMRFS optimisé pour S3 est désormais activé par défaut, ce qui améliore les performances d'écriture. Pour plus d'informations, consultez [Utilisez le transmetteur EMRFS optimisé pour S3](#).
- Hive
 - Rétroporté [HIVE-16686](#).
- Glue avec Spark et Hive
 - Dans la version EMR 5.20.0 ou ultérieure, l'élagage parallèle des partitions est activé automatiquement pour Spark et Hive lorsque AWS Glue Data Catalog est utilisé comme métastore. Cette modification réduit considérablement le temps de planification des requêtes en exécutant plusieurs requêtes en parallèle pour récupérer des partitions. Le nombre total de segments pouvant être exécutés simultanément est compris entre 1 et 10. La valeur par défaut est 5, ce qui est recommandé. Vous pouvez le modifier en spécifiant la propriété `aws.glue.partition.num.segments` dans la classification de configuration `hive-site`. En cas de limitation, vous pouvez désactiver la fonctionnalité en remplaçant la valeur par 1. Pour en savoir plus, consultez [Structure d'un segment AWS Glue](#).

Problèmes connus

- Hue (corrigé dans la EMR version 5.24.0 d'Amazon)
 - Hue exécuté sur Amazon EMR ne prend pas en charge Solr. À compter de la EMR version 5.20.0 d'Amazon, un problème de configuration incorrecte entraîne l'activation de Solr et un message d'erreur inoffensif similaire au suivant s'affiche :

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

Pour empêcher l'affichage du message d'erreur de Solr, procédez comme suit :

1. Connectez-vous à la ligne de commande du nœud principal à l'aide deSSH.
2. Utilisez un éditeur de texte pour ouvrir le fichier `hue.ini`. Par exemple :

```
sudo vim /etc/hue/conf/hue.ini
```

3. Recherchez le terme `appblacklist` et modifiez la ligne comme suit :

```
appblacklist = search
```

4. Enregistrez vos modifications et redémarrez Hue comme indiqué dans l'exemple suivant :

```
sudo stop hue; sudo start hue
```

- Tez
 - Ce problème a été résolu dans Amazon EMR 5.22.0.

Lorsque vous vous connectez à l'interface utilisateur de Tez sur `http://MasterDNS:8080/tez-ui` via une SSH connexion au nœud principal du cluster, l'erreur « L'opération de l'adaptateur a échoué - Le serveur Timeline (ATS) est hors de portée. Soit il est en panne, soit il CORS est « non activé » s'affiche, soit les tâches affichent « N/A » de manière inattendue.

Cela est dû au fait que l'interface utilisateur de Tez envoie des demandes au serveur YARN Timeline en utilisant `localhost` plutôt que le nom d'hôte du nœud principal. Pour contourner le problème, un script est disponible pour être exécuté en tant qu'action ou étape d'amorçage.

Le script met à jour le nom d'hôte dans le fichier `Tez configs .env`. Pour plus d'informations et pour connaître l'emplacement du script, consultez les [Instructions sur l'amorçage](#).

- Dans les EMR versions 5.19.0, 5.20.0 et 5.21.0 d'Amazon, les étiquettes des YARN nœuds sont stockées dans un répertoire. HDFS Dans certaines situations, cela entraîne des retards dans le démarrage des nœuds principaux, ce qui provoque le dépassement du délai du cluster et l'échec du lancement. À partir d'Amazon EMR 5.22.0, ce problème est résolu. YARN Les étiquettes des nœuds sont stockées sur le disque local de chaque nœud du cluster, ce qui permet d'éviter toute dépendance vis-à-vis de HDFS.
- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSH en tant qu'utilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Versions des composants 5.20.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.2.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2.29.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.6.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.8.5-amzn-1	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.8	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.8	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.8	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.4.8	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.4.8	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	2.3.4-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.4-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.4-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.4-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	4.3.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.3.1	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur

Composant	Version	Description
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.214	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.214	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.4.0	Clients de ligne de commande Spark.
spark-history-server	2.4.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.4.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.4.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.12.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.9.1	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.20.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.20.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.

Classifications	Description
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.

Classifications	Description
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.

Classifications	Description
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.

Classifications	Description
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.

Classifications	Description
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.

Classifications	Description
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.19.1

Versions d'application 5.19.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
Kit AWS SDK pour Java	1,1,433	1,1,433	1,11,393	1,11,393
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.1	1.6.1	1.6.0	1.6.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.7	1.4.7
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.5	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.9.4	0.8.1	0.8.1
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.0	1.3.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.212	0.210	0.210
Spark	2.3.2	2.3.2	2.3.2	2.3.2
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.11.0	1.11.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.8.0
ZooKeeper	3.4.13	3.4.13	3.4.12	3.4.12

Notes de mise à jour 5.19.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.19.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent

généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.1.0	EMRConnecteur S3Select
<code>emrfs</code>	2.28.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.6.1	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.8.5-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.8.5-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.7	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.7	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.7	HBaseclient en ligne de commande.
hbase-rest-server	1.4.7	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.7	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hcatalog-client	2.3.3-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.3-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark

Composant	Version	Description
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.3.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.212	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.212	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.2	Clients de ligne de commande Spark.
spark-history-server	2.3.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.

Composant	Version	Description
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.19.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.19.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.

Classifications	Description
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.

Classifications	Description
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.

Classifications	Description
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.

Classifications	Description
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.

Classifications	Description
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.

Classifications	Description
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.19.0

Versions d'application 5.19.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
Kit AWS SDK pour Java	1,1,433	1,11,393	1,11,393	1,1336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.1	1.6.0	1.6.0	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.7	1.4.6

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.5	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.9.4	0.8.1	0.8.1	0.8.1
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.3.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.212	0.210	0.210	0.206
Spark	2.3.2	2.3.2	2.3.2	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.11.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.8.0	0.7.3
ZooKeeper	3.4.13	3.4.12	3.4.12	3.4.12

Notes de mise à jour 5.19.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.19.0 d'Amazon. Les modifications ont été apportées à la version 5.18.0.

Date de parution initiale : 7 novembre 2018

Date de la dernière mise à jour : 19 novembre 2018

Mises à niveau

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0,9.4
- MXNet1.3.0
- Presto 0.212
- TensorFlow 1,11,0
- Zookeeper 3.4.13
- AWS SDK for Java 1,1,433

Nouvelles fonctionnalités

- (19 novembre 2018) EMR Notebooks est un environnement géré basé sur Jupyter Notebook. Il prend en charge les noyaux magiques Spark pour PySpark, SparkSQL, Spark R et Scala. EMR Les blocs-notes peuvent être utilisés avec des clusters créés à l'aide des EMR versions 5.18.0 et ultérieures d'Amazon. Pour plus d'informations, consultez la section [Utilisation EMR des blocs-notes](#) dans le guide EMR de gestion Amazon.

- Le valideur EMRFS optimisé pour S3 est disponible lors de l'écriture de fichiers Parquet à l'aide de Spark et. EMRFS Ce validateur améliore les performances d'écriture. Pour plus d'informations, consultez [Utilisez le transmetteur EMRFS optimisé pour S3](#).

Modifications, améliorations et problèmes résolus

- YARN
 - Modification de la logique qui limite le processus principal de l'application à l'exécution sur les nœuds principaux. Cette fonctionnalité utilise désormais la fonctionnalité et les propriétés des étiquettes de YARN nœuds dans les classifications `capacity-scheduler` de configuration `yarn-site` et. Pour plus d'informations, veuillez consulter <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.
- Amazon Linux par défaut AMI pour Amazon EMR
 - `ruby18`, `php56` et `gcc48` ne sont plus installés par défaut. Ils peuvent être installés si vous le souhaitez à l'aide de `yum`.
 - Le gem `ruby aws-sdk` n'est plus installé par défaut. Il peut être installé en utilisant `gem install aws-sdk`, si vous le souhaitez. Des composants spécifiques peuvent également être installés. Par exemple, `gem install aws-sdk-s3`.

Problèmes connus

- EMRBlocs-notes : dans certains cas, lorsque plusieurs éditeurs de blocs-notes sont ouverts, l'éditeur de bloc-notes peut sembler incapable de se connecter au cluster. Dans ce cas, effacez les cookies du navigateur, puis rouvrez les éditeurs de bloc-notes.
- CloudWatch ContainerPending Mise à l'échelle métrique et automatique — (Corrigé dans la version 5.20.0) Amazon EMR peut émettre une valeur négative pour `ContainerPending`. Si `ContainerPending` est utilisée dans une règle de mise à l'échelle automatique, la mise à l'échelle automatique ne se comporte pas comme prévu. Évitez d'utiliser `ContainerPending` avec la mise à l'échelle automatique.
- Dans les EMR versions 5.19.0, 5.20.0 et 5.21.0 d'Amazon, les étiquettes des YARN nœuds sont stockées dans un répertoire. HDFS Dans certaines situations, cela entraîne des retards dans le démarrage des nœuds principaux, ce qui provoque le dépassement du délai du cluster et l'échec du lancement. À partir d'Amazon EMR 5.22.0, ce problème est résolu. YARN Les étiquettes des nœuds sont stockées sur le disque local de chaque nœud du cluster, ce qui permet d'éviter toute dépendance vis-à-vis de HDFS.

Versions des composants 5.19.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.2.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.1.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2.28.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.6.1	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.5-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.5-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.8.5-amzn-0	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.5-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.7	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.7	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.7	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.4.7	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.4.7	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	2.3.3-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.3-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.9.4	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.3.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur

Composant	Version	Description
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.212	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.212	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.2	Clients de ligne de commande Spark.
spark-history-server	2.3.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.11.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.13	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.13	ZooKeeper client en ligne de commande.

Classifications des configurations 5.19.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.19.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.

Classifications	Description
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.

Classifications	Description
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.

Classifications	Description
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.

Classifications	Description
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier <code>password-authenticator.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.

Classifications	Description
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-memory	Modifiez les valeurs dans le fichier memory.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.

Classifications	Description
presto-connector-tpcds	Modifiez les valeurs dans le fichier tpcds.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.18.1

Versions d'application 5.18.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
Kit AWS SDK pour Java	1,11,393	1,11,393	1,1336	1,1336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.0	1.6.0	1.5.2	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.7	1.4.6	1.4.6
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.210	0.210	0.206	0.206
Spark	2.3.2	2.3.2	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.8.0	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

Notes de mise à jour 5.18.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.18.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent

généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.5.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.1.0	EMRConnecteur S3Select
<code>emrfs</code>	2.27.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.6.0	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.4-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.4-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.8.4-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.4-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.7	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.7	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.7	HBaseclient en ligne de commande.
hbase-rest-server	1.4.7	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.7	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hcatalog-client	2.3.3-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.3-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark

Composant	Version	Description
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.2.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.210	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.210	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.2	Clients de ligne de commande Spark.
spark-history-server	2.3.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.9.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.

Composant	Version	Description
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.12	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.12	ZooKeeper client en ligne de commande.

Classifications des configurations 5.18.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.18.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.

Classifications	Description
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.

Classifications	Description
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.

Classifications	Description
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.

Classifications	Description
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.

Classifications	Description
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark

Classifications	Description
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.18.0

Versions d'application 5.18.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
Kit AWS SDK pour Java	1,11,393	1,1336	1,1336	1,1336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.6.0	1.5.2	1.5.2	1.5.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.7	1.4.6	1.4.6	1.4.6

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.210	0.206	0.206	0.206
Spark	2.3.2	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	1.9.0
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.8.0	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

Notes de mise à jour 5.18.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.18.0 d'Amazon. Les modifications ont été apportées à la version 5.17.0.

Date de parution initiale : 24 octobre 2018

Mises à niveau

- Flink 1.6.0
- HBase1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

Nouvelles fonctionnalités

- À partir d'Amazon EMR 5.18.0, vous pouvez utiliser le référentiel d'EMRartefacts Amazon pour créer votre code de travail en fonction des versions exactes des bibliothèques et des dépendances disponibles avec des versions spécifiques d'Amazon. EMR Pour plus d'informations, consultez [Vérification des dépendances à l'aide du référentiel d'artefacts d'Amazon EMR](#).

Modifications, améliorations et problèmes résolus

- Hive
 - Ajout de la prise en charge de S3 Select. Pour plus d'informations, consultez [Utilisation de S3 Select avec Hive pour améliorer les performances](#).
- Presto

- Ajout de la prise en charge de [S3 Select](#) Pushdown. Pour plus d'informations, consultez [Utilisation de S3 Select Pushdown avec Presto pour améliorer les performances](#).
- Spark
 - La configuration par défaut de log4j pour Spark a été modifiée pour enregistrer les journaux des conteneurs toutes les heures pour les tâches de streaming Spark. Cela permet d'éviter la suppression des journaux pour les tâches de streaming Spark de longue durée.

Versions des composants 5.18.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.5.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.1.0	EMRConnecteur S3Select
emrfs	2.27.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.6.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.4-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».

Composant	Version	Description
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.4-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.8.4-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.4-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	1.4.7	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.7	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.7	HBaseclient en ligne de commande.
hbase-rest-server	1.4.7	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.7	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.3-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.3-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-2	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.3-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
nginx	1.12.1	nginx [engine x] est HTTP un serveur proxy inverse
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.2.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda

Composant	Version	Description
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.210	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.210	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.2	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.3.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.9.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.8.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.12	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.12	ZooKeeper client en ligne de commande.

Classifications des configurations 5.18.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.18.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.

Classifications	Description
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.

Classifications	Description
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.17.2

Versions d'application 5.17.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
Kit AWS SDK pour Java	1,1336	1,1336	1,1336	1,1336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.2	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.6	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.206	0.206	0.203
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	1.9.0	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

Notes de mise à jour 5.17.2

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.17.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.5.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.0.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2.26.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.5.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.4-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.4-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.8.4-amzn-1	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.4-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.6	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.6	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.6	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.4.6	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.4.6	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	2.3.3-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.3-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.2.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.206	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.206	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.1	Clients de ligne de commande Spark.
spark-history-server	2.3.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	1.9.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.12	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.12	ZooKeeper client en ligne de commande.

Classifications des configurations 5.17.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.17.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.

Classifications	Description
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog <code>proto-hive-site</code> fichier <code>.xml</code> .
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat <code>log4j2.properties</code> de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier <code>webhcat-site.xml</code> de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier <code>beeline-log4j2.properties</code> de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier <code>oozie-log4j.properties</code> d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier <code>oozie-site.xml</code> d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.

Classifications	Description
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.

Classifications	Description
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.17.1

Versions d'application 5.17.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
Kit AWS SDK pour Java	1,1336	1,1336	1,1336	1,1336
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.2	1.5.0	1.5.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.6	1.4.4	1.4.4

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.4
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0,5,0	0,5,0	0,5,0	0,5,0
MXNet	1.2.0	1.2.0	1.2.0	1.2.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.14.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.206	0.203	0.203
Spark	2.3.1	2.3.1	2.3.1	2.3.1
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	1.9.0	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

Notes de mise à jour 5.17.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.17.1 d'Amazon. Les modifications ont été apportées à la version 5.17.0.

Date de parution initiale : 18 juillet 2019

Modifications, améliorations et problèmes résolus

- Mise à jour de la version par défaut d'Amazon Linux AMI pour Amazon EMR afin d'inclure d'importantes mises à jour de sécurité du noyau Linux, notamment le problème de TCP SACK déni de service ([AWS-2019-005](#)).

Versions des composants 5.17.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.5.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
emr-s3-select	1.0.0	EMRConnecteur S3Select
emrfs	2.26.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.5.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.4-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.4-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.8.4-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.4-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.6	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.6	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.6	HBaseclient en ligne de commande.
hbase-rest-server	1.4.6	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.6	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.3-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.

Composant	Version	Description
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.3-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	REST interface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.2.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.

Composant	Version	Description
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.206	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.206	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.

Composant	Version	Description
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.1	Clients de ligne de commande Spark.
spark-history-server	2.3.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tensorflow	1.9.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.12	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.12	ZooKeeper client en ligne de commande.

Classifications des configurations 5.17.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.17.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.

Classifications	Description
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2

Classifications	Description
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.

Classifications	Description
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.17.0

Versions d'application 5.17.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
Kit AWS SDK pour Java	1,1336	1,1336	1,1336	1,1333
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.2	1.5.0	1.5.0	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.6	1.4.4	1.4.4	1.4.4
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.4	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0,5,0	0,5,0	0,5,0	0,4,0
MXNet	1.2.0	1.2.0	1.2.0	1.1.0

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.14.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.206	0.203	0.203	0.194
Spark	2.3.1	2.3.1	2.3.1	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	1.9.0	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

Notes de mise à jour 5.17.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.17.0 d'Amazon. Les modifications ont été apportées à la version 5.16.0.

Date de parution initiale : 30 août 2018

Mises à niveau

- Flink 1.5.2
- HBase1.4.6
- Presto 0.206

Nouvelles fonctionnalités

- Ajout de la prise en charge de Tensorflow. Pour plus d'informations, consultez [TensorFlow](#).

Modifications, améliorations et problèmes résolus

- JupyterHub
 - Ajout de la prise en charge de la persistance des blocs-notes dans Amazon S3. Pour plus d'informations, consultez [Configuration de la persistance pour les blocs-notes dans Amazon S3](#).
- Spark
 - Ajout de la prise en charge de [S3 Select](#). Pour plus d'informations, consultez [Utilisation de S3 Select avec Spark pour améliorer les performances des requêtes](#).
- Résolution des problèmes liés aux métriques Cloudwatch et à la fonctionnalité de dimensionnement automatique dans les EMR versions 5.14.0, 5.15.0 ou 5.16.0 d'Amazon.

Problèmes connus

- Lorsque vous créez un cluster activé pour Kerberos avec Livy installé, Livy échoue avec un message d'erreur indiquant que l'authentification simple n'est pas activée. Le redémarrage du serveur Livy résout le problème. Pour contourner le problème, ajoutez une étape lors de la création du cluster qui exécute `sudo restart livy-server` sur le nœud primaire.
- Si vous utilisez un Amazon Linux personnalisé AMI basé sur un Amazon Linux AMI dont la date de création est le 11/08/2018, le serveur Oozie ne démarre pas. Si vous utilisez Oozie, créez une personnalisation AMI basée sur un AMI identifiant Amazon Linux avec une date de création différente. Vous pouvez utiliser la AWS CLI commande suivante pour renvoyer une liste d'images IDs pour tous les HVM Amazon Linux AMIs avec une version 2018.03, ainsi que la date de sortie, afin de pouvoir choisir un Amazon Linux approprié AMI comme base. MyRegion Remplacez-le par l'identifiant de votre région, tel que us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`][?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

Versions des composants 5.17.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.5.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emr-s3-select</code>	1.0.0	EMRConnecteur S3Select

Composant	Version	Description
emrfs	2.26.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.5.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.4-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.4-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.8.4-amzn-1	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.4-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.6	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.6	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.6	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.4.6	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.4.6	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	2.3.3-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.3-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.2.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.206	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.206	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.1	Clients de ligne de commande Spark.
spark-history-server	2.3.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tensorflow	1.9.0	TensorFlow bibliothèque de logiciels open source pour le calcul numérique de haute performance.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.12	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.12	ZooKeeper client en ligne de commande.

Classifications des configurations 5.17.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.17.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.

Classifications	Description
container-log4j	Modifiez les valeurs dans le fichier container-log4j.properties YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog <code>proto-hive-site</code> fichier <code>.xml</code> .
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat <code>log4j2.properties</code> de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier <code>webhcat-site.xml</code> de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier <code>beeline-log4j2.properties</code> de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
jupyter-notebook-conf	Modifiez les valeurs dans le fichier <code>jupyter_notebook_config.py</code> de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .
jupyter-s3-conf	Configuration de la persistance S3 de bloc-notes Jupyter.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .
mapred-env	Modifiez les valeurs dans l'environnement de l'MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier <code>oozie-log4j.properties</code> d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier <code>oozie-site.xml</code> d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.

Classifications	Description
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.

Classifications	Description
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.16.1

Versions d'application 5.16.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
Kit AWS SDK pour Java	1,1336	1,1336	1,1333	1,1333
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.0	1.5.0	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.4	1.4.4

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
HCatalog	2.3.3	2.3.3	2.3.3	2.3.3
Hadoop	2.8.4	2.8.4	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.3
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.2.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0,5,0	0,5,0	0,4,0	0,4,0
MXNet	1.2.0	1.2.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	5.0.0
Phoenix	4.14.0	4.14.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.203	0.203	0.194	0.194
Spark	2.3.1	2.3.1	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.12

Notes de mise à jour 5.16.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.16.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.1.0	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.25.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.5.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.4-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.4-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.8.4-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.4-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.8.4-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.4	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.4	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.4	HBaseclient en ligne de commande.
hbase-rest-server	1.4.4	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.4	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.3-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.

Composant	Version	Description
hcatalog-webhcat-server	2.3.3-amzn-1	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.3-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.2.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.

Composant	Version	Description
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.203	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.203	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.

Composant	Version	Description
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.1	Clients de ligne de commande Spark.
spark-history-server	2.3.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.12	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.12	ZooKeeper client en ligne de commande.

Classifications des configurations 5.16.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.16.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.

Classifications	Description
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.

Classifications	Description
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.

Classifications	Description
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.

Classifications	Description
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.

Classifications	Description
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.16.0

Versions d'application 5.16.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
Kit AWS SDK pour Java	1,1336	1,1333	1,1333	1,11,297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.5.0	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.4	1.4.2
HCatalog	2.3.3	2.3.3	2.3.3	2.3.2
Hadoop	2.8.4	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.3	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.2.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0,5,0	0,4,0	0,4,0	0,4,0
MXNet	1.2.0	1.1.0	1.1.0	1.1.0

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	5.0.0	4.3.0
Phoenix	4.14.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.203	0.194	0.194	0.194
Spark	2.3.1	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.12	3.4.10

Notes de mise à jour 5.16.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.16.0 d'Amazon. Les modifications ont été apportées à la version 5.15.0.

Date de parution initiale : 19 juillet 2018

Mises à niveau

- Hadoop 2.8.4
- Flink 1.5.0
- Livy 0.5.0
- MXNet1.2.0

- Phoenix 4.14.0
- Presto 0.203
- Spark 2.3.1
- AWS SDK for Java 1,1336
- CUDA9,2
- Pilote Redshift 1.2.15.1025 JDBC

Modifications, améliorations et problèmes résolus

- HBase
 - [Rétroporté -20723 HBASE](#)
- Presto
 - Modifications de configuration pour prendre en charge LDAP l'authentification. Pour plus d'informations, consultez [Utilisation de LDAP l'authentification pour Presto sur Amazon EMR](#).
- Spark
 - [Apache Spark version 2.3.1, disponible à partir de la EMR version 5.16.0 d'Amazon, adresses CVE-2018-8024 et -2018-1334. CVE](#) Nous vous recommandons de migrer les versions antérieures de Spark vers la version 2.3.1 ou ultérieure.

Problèmes connus

- Cette version ne prend pas en charge les types d'instance c1.medium ou m1.small. Les clusters utilisant l'un ou l'autre de ces types d'instances ne démarrent pas. Pour contourner le problème, spécifiez un type d'instance différent ou utilisez une version différente.
- Lorsque vous créez un cluster activé pour Kerberos avec Livy installé, Livy échoue avec un message d'erreur indiquant que l'authentification simple n'est pas activée. Le redémarrage du serveur Livy résout le problème. Pour contourner le problème, ajoutez une étape lors de la création du cluster qui exécute `sudo restart livy-server` sur le nœud primaire.
- Après le redémarrage du nœud principal ou le redémarrage du contrôleur d'instance, les CloudWatch métriques ne seront pas collectées et la fonctionnalité de dimensionnement automatique ne sera pas disponible dans les EMR versions 5.14.0, 5.15.0 ou 5.16.0 d'Amazon. Ce problème est résolu dans Amazon EMR 5.17.0.

Versions des composants 5.16.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.1.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.25.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
flink-client	1.5.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.4-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.4-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.8.4-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.8.4-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.4-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.4	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.4	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.4	HBaseclient en ligne de commande.
hbase-rest-server	1.4.4	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.4.4	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.3-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.3-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.5.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.2.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.2.88	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,14,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,14,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.203	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.203	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.1	Clients de ligne de commande Spark.
spark-history-server	2.3.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.12	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.12	ZooKeeper client en ligne de commande.

Classifications des configurations 5.16.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.16.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .

Classifications	Description
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.

Classifications	Description
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.

Classifications	Description
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.

Classifications	Description
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-password-authenticator	Modifiez les valeurs dans le fichier password-authenticator.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.

Classifications	Description
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.15.1

Versions d'application 5.15.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)

- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
Kit AWS SDK pour Java	1,1333	1,1333	1,11,297	1,11,297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.4	1.4.2	1.4.2
HCatalog	2.3.3	2.3.3	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.3	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.2.0	4.2.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	5.0.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.12	3.4.10	3.4.10

Notes de mise à jour 5.15.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

⚠ Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.15.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-s3-dist-cp	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.24.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.4.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.3-amzn-1	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.8.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.4	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.4	Service pour desservir une ou plusieurs HBase régions.

Composant	Version	Description
hbase-client	1.4.4	HBaseclient en ligne de commande.
hbase-rest-server	1.4.4	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.4	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.3-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.3-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.3-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server2	2.3.3-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.1.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur

Composant	Version	Description
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.194	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.194	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.0	Clients de ligne de commande Spark.
spark-history-server	2.3.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.12	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.12	ZooKeeper client en ligne de commande.

Classifications des configurations 5.15.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.15.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.

Classifications	Description
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.

Classifications	Description
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.

Classifications	Description
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.

Classifications	Description
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier <code>jupyterhub_config.py</code> .
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier <code>config.json</code> de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier <code>livy.conf</code> de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy <code>log4j.properties</code> .
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier <code>mapred-site.xml</code> de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier <code>oozie-log4j.properties</code> d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier <code>oozie-site.xml</code> d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.

Classifications	Description
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.15.0

Versions d'application 5.15.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
Kit AWS SDK pour Java	1,1333	1,11,297	1,11,297	1,11,297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.4	1.4.2	1.4.2	1.4.2
HCatalog	2.3.3	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.3	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
Hue	4.2.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	0.8.1
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.1.0	1.1.0	1.1.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	5.0.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.7
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.12	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.15.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.15.0 d'Amazon. Les modifications ont été apportées à la version 5.14.0.

Date de parution initiale : 21 juin 2018

Mises à niveau

- Mise à niveau HBase vers la version 1.4.4
- Mise à niveau de Hive vers la version 2.3.3
- Mise à niveau de Hue vers la version 4.2.0
- Mise à niveau d'Oozie vers la version 5.0.0
- Mise à niveau de Zookeeper vers la version 3.4.12
- Mise à niveau AWS SDK vers la version 1.11.333

Modifications, améliorations et problèmes résolus

- Hive
 - [Rétroporté -18069 HIVE](#)
- Hue
 - Hue a été mis à jour pour s'authentifier correctement auprès de Livy lorsque Kerberos est activé. Livy est désormais pris en charge lors de l'utilisation de Kerberos avec Amazon. EMR
- JupyterHub
 - Mis à jour JupyterHub afin qu'Amazon EMR installe les bibliothèques LDAP clientes par défaut.
 - Correction d'une erreur dans le script qui génère des certificats auto-signés.

Problèmes connus

- Cette version ne prend pas en charge les types d'instance c1.medium ou m1.small. Les clusters utilisant l'un ou l'autre de ces types d'instances ne démarrent pas. Pour contourner le problème, spécifiez un type d'instance différent ou utilisez une version différente.
- Après le redémarrage du nœud principal ou le redémarrage du contrôleur d'instance, les CloudWatch métriques ne seront pas collectées et la fonctionnalité de dimensionnement automatique ne sera pas disponible dans les EMR versions 5.14.0, 5.15.0 ou 5.16.0 d'Amazon. Ce problème est résolu dans Amazon EMR 5.17.0.

Versions des composants 5.15.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.24.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
flink-client	1.4.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.8.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.8.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.4	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.4	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.4	HBaseclient en ligne de commande.
hbase-rest-server	1.4.4	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.4.4	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.3-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.3-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.3-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.3-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.3-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.3-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.3-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.2.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.1.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	5.0.0	Client de ligne de commande Oozie.
oozie-server	5.0.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.194	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.194	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.0	Clients de ligne de commande Spark.
spark-history-server	2.3.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.12	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.12	ZooKeeper client en ligne de commande.

Classifications des configurations 5.15.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.15.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .

Classifications	Description
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.

Classifications	Description
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.

Classifications	Description
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.

Classifications	Description
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.

Classifications	Description
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.14.2

Versions d'application 5.14.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
Kit AWS SDK pour Java	1,11,297	1,11,297	1,11,297	1,11,297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.2	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.2
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	0.8.1	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.1.0	1.1.0	1.0.0

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.7	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.14.2

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.14.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.23.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
flink-client	1.4.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.8.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.8.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.2	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.4.2	HBaseclient en ligne de commande.
hbase-rest-server	1.4.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.4.2	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.2-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.1.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.1.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.194	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.194	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.0	Clients de ligne de commande Spark.
spark-history-server	2.3.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.14.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.14.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .

Classifications	Description
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.

Classifications	Description
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.

Classifications	Description
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.

Classifications	Description
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.

Classifications	Description
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.14.1

Versions d'application 5.14.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
Kit AWS SDK pour Java	1,11,297	1,11,297	1,11,297	1,11,297
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.2	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.2
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	0.8.1	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.1.0	1.0.0	1.0.0

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.194
Spark	2.3.0	2.3.0	2.3.0	2.3.0
Sqoop	1.4.7	1.4.7	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.14.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.14.1 d'Amazon. Les modifications ont été apportées à la version 5.14.0.

Date de parution initiale : 17 octobre 2018

Mise à jour de la valeur par défaut AMI pour Amazon EMR afin de corriger les failles de sécurité potentielles.

Versions des composants 5.14.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à

Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.23.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.4.2	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.8.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.4.2	HBaseclient en ligne de commande.
hbase-rest-server	1.4.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.4.2	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.2-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.1.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.1.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.194	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.194	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.0	Clients de ligne de commande Spark.
spark-history-server	2.3.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.14.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.14.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .

Classifications	Description
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.

Classifications	Description
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.

Classifications	Description
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.

Classifications	Description
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.

Classifications	Description
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.14.0

Versions d'application 5.14.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
Kit AWS SDK pour Java	1,11,297	1,11,297	1,11,297	1,11,267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.2	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.2	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	0.8.1	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.1.0	1.0.0	1.0.0	1.0.0

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.194	0.188
Spark	2.3.0	2.3.0	2.3.0	2.2.1
Sqoop	1.4.7	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.14.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.14.0 d'Amazon. Les modifications ont été apportées à la version 5.13.0.

Date de parution initiale : 4 juin 2018

Mises à niveau

- Mise à niveau d'Apache Flink vers la version 1.4.2
- Mise à niveau d'Apache MXnet vers la version 1.1.0
- Mise à niveau d'Apache Sqoop vers la version 1.4.7

Nouvelles fonctionnalités

- JupyterHub Support ajouté. Pour plus d'informations, consultez [JupyterHub](#).

Modifications, améliorations et problèmes résolus

- EMRFS
 - La userAgent chaîne des demandes adressées à Amazon S3 a été mise à jour pour contenir les informations d'utilisateur et de groupe du principal invocateur. Cela peut être utilisé avec AWS CloudTrail les journaux pour un suivi plus complet des demandes.
- HBase
 - Inclus [HBASE-20447](#), qui résout un problème susceptible de provoquer des problèmes de cache, en particulier dans le cas de régions divisées.
- MXnet
 - Ajout des bibliothèques OpenCV.
- Spark
 - Lorsque Spark écrit des fichiers Parquet sur un emplacement Amazon S3 en utilisant EMRFS, l'FileOutputCommitter algorithme a été mis à jour pour utiliser la version 2 au lieu de la version 1. Cela réduit le nombre de renommages, ce qui améliore les performances de l'application. Cette modification n'affecte pas :
 - Les applications autres que Spark.
 - Applications qui écrivent sur d'autres systèmes de fichiers, tels que HDFS (qui utilisent toujours la version 1 de FileOutputCommitter).
 - Applications utilisant d'autres formats de sortie, tels que le texte ou le csv, qui utilisent déjà l'écriture EMRFS directe.

Problèmes connus

- JupyterHub
 - L'utilisation de classifications de configuration pour configurer JupyterHub des blocs-notes Jupyter individuels lors de la création d'un cluster n'est pas prise en charge. Modifiez manuellement le fichier jupyterhub_config.py et les fichiers jupyter_notebook_config.py pour chaque utilisateur. Pour plus d'informations, consultez [Configuration JupyterHub](#).

- JupyterHub ne démarre pas sur des clusters au sein d'un sous-réseau privé, ce qui entraîne un échec du messageError: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt' . Ce problème est dû à une erreur dans le script qui génère des certificats auto-signés. Utilisez la solution suivante pour générer des certificats auto-signés. Toutes les commandes sont exécutées lorsque vous êtes connecté au nœud primaire.

1. Copiez le script de génération de certificat du conteneur vers le nœud primaire :

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Utilisez un éditeur de texte pour modifier la ligne 23 afin de changer le nom d'hôte public en nom d'hôte local, comme indiqué ci-dessous :

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Exécutez le script pour générer des certificats auto-signés :

```
sudo bash ./gen_self_signed_cert.sh
```

4. Déplacez les fichiers de certificat générés par le script vers le répertoire /etc/jupyter/conf/ :

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

Vous pouvez consulter `tail` le `jupyter.log` fichier pour vérifier qu'il JupyterHub a redémarré et qu'il renvoie un code de réponse 200. Par exemple :

```
tail -f /var/log/jupyter/jupyter.log
```

Vous devriez obtenir une réponse similaire à la suivante :

```
# [I 2018-06-14 18:56:51.356 JupyterHub app:1581] JupyterHub is now running at  
https://:9443/  
# 19:01:51.359 - info: [ConfigProxy] 200 GET /api/routes
```

- Après le redémarrage du nœud principal ou le redémarrage du contrôleur d'instance, les CloudWatch métriques ne seront pas collectées et la fonctionnalité de dimensionnement automatique ne sera pas disponible dans les EMR versions 5.14.0, 5.15.0 ou 5.16.0 d'Amazon. Ce problème est résolu dans Amazon EMR 5.17.0.

Versions des composants 5.14.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.23.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
flink-client	1.4.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.8.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.8.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.4.2	HBaseclient en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.4.2	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.4.2	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	2.3.2-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.2-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	4.1.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
jupyterhub	0.8.1	Serveur multi-utilisateurs pour blocs-notes Jupyter
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.1.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
opencv	3.4.0	Bibliothèque open source Vision par ordinateur
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.194	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.194	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.0	Clients de ligne de commande Spark.
spark-history-server	2.3.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.7	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.14.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.14.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
container-log4j	Modifiez les valeurs dans le fichier <code>container-log4j.properties</code> YARN de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.

Classifications	Description
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.

Classifications	Description
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.

Classifications	Description
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
jupyter-notebook-conf	Modifiez les valeurs dans le fichier jupyter_notebook_config.py de Jupyter Notebook.
jupyter-hub-conf	Modifiez les valeurs dans JupyterHubs le fichier jupyterhub_config.py.

Classifications	Description
jupyter-sparkmagic-conf	Modifiez les valeurs dans le fichier config.json de Sparkmagic.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.

Classifications	Description
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.

Classifications	Description
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.13.1

Versions d'application 5.13.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
Kit AWS SDK pour Java	1,11,297	1,11,297	1,11,267	1,11,267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2,11.8	2,11.8	2,11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.2	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.194	0.188	0.188
Spark	2.3.0	2.3.0	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.13.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

⚠ Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.13.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-s3-dist-cp	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.22.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.4.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.3-amzn-0	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.8.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.4.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.4.2	HBaseclient en ligne de commande.
hbase-rest-server	1.4.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.2	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.2-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.2-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.1.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.0.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.194	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.194	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.0	Clients de ligne de commande Spark.
spark-history-server	2.3.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.13.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.13.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.

Classifications	Description
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.

Classifications	Description
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.

Classifications	Description
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.

Classifications	Description
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.13.0

Versions d'application 5.13.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
Kit AWS SDK pour Java	1,11,297	1,11,267	1,11,267	1,11,267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2,11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.2	1.4.0	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.194	0.188	0.188	0.188
Spark	2.3.0	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.13.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.13.0 d'Amazon. Les modifications ont été apportées à la version 5.12.0.

Mises à niveau

- Mise à niveau de Spark vers la version 2.3.0
- Mise à niveau HBase vers la version 1.4.2
- Mise à niveau vers de Presto vers la version 0.194
- Mise à niveau AWS SDK for Java vers la version 1.11.297

Modifications, améliorations et problèmes résolus

- Hive
 - Rétroporté [HIVE-15436](#). Hive améliorée pour ne APIs renvoyer que les vues.

Problèmes connus

- MXNet ne possède actuellement pas de bibliothèques OpenCV.

Versions des composants 5.13.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.10.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.22.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.4.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.8.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-ftpfs-server	2.8.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.4.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.4.2	HBaseclient en ligne de commande.
hbase-rest-server	1.4.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.2	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.2-amzn-2	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-2	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-2	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.2-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.1.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.0.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.194	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.194	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
r	3.4.1	Projet R pour les calculs statistiques
spark-client	2.3.0	Clients de ligne de commande Spark.
spark-history-server	2.3.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.3.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.3.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Composant	Version	Description
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.13.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.13.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.

Classifications	Description
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.

Classifications	Description
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.

Classifications	Description
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.

Classifications	Description
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.12.3

Versions d'application 5.12.3

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
Kit AWS SDK pour Java	1,11,267	1,11,267	1,11,267	1,11,267
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.4.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.4.0	1.4.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.8.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.1.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	1.0.0

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.13.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.188	0.188
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.12.3

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.12.3

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.9.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.21.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
flink-client	1.4.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.8.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.8.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.4.0	HBaseclient en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.4.0	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.4.0	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.2-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.2-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	4.1.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.0.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.188	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.188	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.1	Clients de ligne de commande Spark.
spark-history-server	2.2.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.12.3

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.12.3

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .

Classifications	Description
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue

Classifications	Description
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.

Classifications	Description
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.

Classifications	Description
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.12.2

Versions d'application 5.12.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
Kit AWS SDK pour Java	1,11,267	1,11,267	1,11,267	1,11,238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.4.0	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.4.0	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.8.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.1.0	4.1.0	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	1.0.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.13.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.188	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.12.2

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.12.2 d'Amazon. Les modifications ont été apportées à la version 5.12.1.

Date de parution initiale : 29 août 2018

Modifications, améliorations et problèmes résolus

- Cette version corrige une vulnérabilité de sécurité potentielle.

Versions des composants 5.12.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-`

composant avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.9.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.21.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.4.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les

Composant	Version	Description
		métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.8.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.4.0	HBaseclient en ligne de commande.
hbase-rest-server	1.4.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.2-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.2-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-1	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.1.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.0.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.

Composant	Version	Description
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.188	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.188	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.1	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.2.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.2.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.12.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.12.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase

Classifications	Description
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.

Classifications	Description
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.12.1

Versions d'application 5.12.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
Kit AWS SDK pour Java	1,11,267	1,11,267	1,11,238	1,11,238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.4.0	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.4.0	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.8.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
Hue	4.1.0	4.1.0	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	1.0.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.13.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.188	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.12.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.12.1 d'Amazon. Les modifications ont été apportées à la version 5.12.0.

Date de parution initiale : 29 mars 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités potentielles.

Versions des composants 5.12.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.9.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.21.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.4.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.8.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.8.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.4.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.4.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.4.0	HBaseclient en ligne de commande.
hbase-rest-server	1.4.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.2-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-1	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.2-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.1.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.0.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.188	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.188	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.1	Clients de ligne de commande Spark.
spark-history-server	2.2.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.2.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.12.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.12.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.

Classifications	Description
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.

Classifications	Description
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.

Classifications	Description
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.

Classifications	Description
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.

Classifications	Description
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.12.0

Versions d'application 5.12.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
Kit AWS SDK pour Java	1,11,267	1,11,238	1,11,238	1,11,238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.4.0	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.4.0	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.8.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.1.0	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	1.0.0	0.12.0	0.12.0	0.12.0

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.13.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.188	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.12.0

Les notes de publication suivantes contiennent des informations relatives à la EMR version 5.12.0 d'Amazon. Les modifications ont été apportées à la version 5.11.1.

Mises à niveau

- AWS SDK pour Java 1.11.238 ⇒ 1.11.267. Pour plus d'informations, consultez la section [AWS SDK for Java Change Log](#) on GitHub.
- Hadoop 2.7.3 ⇒ 2.8.3. Pour plus d'informations, consultez [Versions d'Apache Hadoop](#).
- Flink 1.3.2 ⇒ 1.4.0. Pour plus d'informations, consultez [Annonce de publication d'Apache Flink 1.4.0](#).
- HBase 1.3.1 ⇒ 1.4.0. Pour plus d'informations, consultez l'[annonce HBase de sortie](#).

- Hue 4.0.1 ⇒ 4.1.0. Pour plus d'informations, veuillez consulter les [Notes de mise à jour](#).
- MxNet 0,12,0 ⇒ 10,0. Pour plus d'informations, consultez la section [MXNetChange Log](#) on GitHub.
- Presto 0.187 ⇒ 0.188. Pour plus d'informations, veuillez consulter les [Notes de mise à jour](#).

Modifications, améliorations et problèmes résolus

- Hadoop
 - La propriété `yarn.resourcemanager.decommissioning.timeout` a été remplacée par `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`. Vous pouvez utiliser cette propriété pour personnaliser la réduction de la taille du cluster. Pour plus d'informations, consultez [Cluster Scale-Down dans le](#) guide de EMRgestion Amazon.
 - Hadoop CLI a ajouté l'option à la commande `cp` (copy), qui spécifie la copie directe. Vous pouvez l'utiliser pour éviter de créer un fichier `.COPYING` intermédiaire, ce qui accélère la copie de données entre Amazon S3. Pour plus d'informations, consultez [HADOOP-12384](#).
- Pig
 - Ajout de la classification de configuration `pig-env`, qui simplifie la configuration des propriétés de l'environnement Pig. Pour plus d'informations, consultez [Configuration des applications](#).
- Presto
 - Ajout de la classification de configuration, `presto-connector-redshift` que vous pouvez utiliser pour configurer des valeurs dans le fichier de configuration `redshift.properties` de Presto. Pour plus d'informations, consultez [Connecteur Redshift](#) dans la documentation Presto, et [Configuration des applications](#).
 - Le support Presto pour EMRFS a été ajouté et constitue la configuration par défaut. Les EMR versions précédentes d'Amazon utilisaient `PrestOS3FileSystem`, qui était la seule option. Pour plus d'informations, consultez [EMRFS configuration de PrestOS3 FileSystem](#).

Note

Si vous interrogez des données sous-jacentes dans Amazon S3 avec Amazon EMR version 5.12.0, des erreurs Presto peuvent se produire. Cela est dû au fait que Presto ne parvient pas à récupérer les valeurs de classification de configuration depuis `emrfs-site.xml`. Pour contourner le problème, créez un sous-répertoire `emrfs` sous `usr/lib/presto/plugin/hive-hadoop2/` et un lien symbolique dans `usr/lib/presto/plugin/hive-hadoop2/emrfs` vers le fichier `/usr/share/aws/emr/`

`emrfs/conf/emrfs-site.xml` existant. Redémarrez ensuite le processus `presto-server` (`sudo presto-server stop` suivi de `sudo presto-server start`).

- Spark
 - Rétroporté [SPARK-22036 : BigDecimal la multiplication renvoie parfois zéro](#).

Problèmes connus

- MXNet n'inclut pas les bibliothèques OpenCV.
- SparkR n'est pas disponible pour les clusters créés à l'aide d'un custom car AMI R n'est pas installé par défaut sur les nœuds du cluster.

Versions des composants 5.12.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-goodies	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.9.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.21.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.4.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.8.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».

Composant	Version	Description
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.8.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.8.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.8.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.

Composant	Version	Description
hbase-hmaster	1.4.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.4.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.4.0	HBaseclient en ligne de commande.
hbase-rest-server	1.4.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.4.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.2-amzn-1	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-1	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-1	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.2-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.1.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	1.0.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.1.85	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
phoenix-library	4,13,0- -1,4 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,13,0- -1,4 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.188	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.188	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.1	Clients de ligne de commande Spark.
spark-history-server	2.2.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.12.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.12.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.

Classifications	Description
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.

Classifications	Description
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.

Classifications	Description
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-env	Modifiez les valeurs dans l'environnement Pig.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.

Classifications	Description
presto-connector-redshift	Modifiez les valeurs dans le fichier redshift.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.

Classifications	Description
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.11.4

Versions d'application 5.11.4

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
Kit AWS SDK pour Java	1,11,238	1,11,238	1,11,238	1,11,238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.11.4

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.11.4

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.8.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.20.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
flink-client	1.3.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-6	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-6	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-6	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.7.3-amzn-6	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	2.3.2-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.2-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	4.0.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	0.12.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.0.176	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.187	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.187	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.1	Clients de ligne de commande Spark.
spark-history-server	2.2.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.11.4

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.11.4

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .

Classifications	Description
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue

Classifications	Description
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.

Classifications	Description
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.

Classifications	Description
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.11.3

Versions d'application 5.11.3

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
Kit AWS SDK pour Java	1,11,238	1,11,238	1,11,238	1,11,238
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
HCatalog	2.3.2	2.3.2	2.3.2	2.3.2
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.2
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.11.3

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.11.3 d'Amazon. Les modifications ont été apportées à la version 5.11.2.

Date de parution initiale : 18 juillet 2019

Modifications, améliorations et problèmes résolus

- Mise à jour de la version par défaut d'Amazon Linux AMI pour Amazon EMR afin d'inclure d'importantes mises à jour de sécurité du noyau Linux, notamment le problème de TCP SACK déni de service ([AWS-2019-005](#)).

Versions des composants 5.11.3

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.8.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.20.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-6	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-6	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-6	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-6	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.2-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.2-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.0.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	0.12.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.

Composant	Version	Description
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.0.176	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.187	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.187	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.1	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.2.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.11.3

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.11.3

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase

Classifications	Description
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.

Classifications	Description
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.11.2

Versions d'application 5.11.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
Kit AWS SDK pour Java	1,11,238	1,11,238	1,11,238	1,11,221
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.2	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.2	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.1	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.11.2

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.11.2 d'Amazon. Les modifications ont été apportées à la version 5.11.1.

Date de parution initiale : 29 août 2018

Modifications, améliorations et problèmes résolus

- Cette version corrige une vulnérabilité de sécurité potentielle.

Versions des composants 5.11.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-s3-dist-cp	2.8.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.20.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-6	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-6	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.7.3-amzn-6	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-6	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.2-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.2-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.0.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	0.12.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.0.176	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.187	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.187	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.1	Clients de ligne de commande Spark.
spark-history-server	2.2.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.11.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.11.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.

Classifications	Description
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.

Classifications	Description
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.

Classifications	Description
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.

Classifications	Description
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Classifications	Description
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.11.1

Versions d'application 5.11.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
Kit AWS SDK pour Java	1,11,238	1,11,238	1,11,221	1,11,221
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.2	2.3.2	2.3.1	2.3.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.2	2.3.1	2.3.1
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	0.12.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
Presto	0.187	0.187	0.187	0.187
Spark	2.2.1	2.2.1	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.3
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.11.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.11.1 d'Amazon. Les modifications concernent la version EMR 5.8.0 d'Amazon.

Date de parution initiale : 22 janvier 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités associées à l'exécution spéculative (CVE-2017-5715, -2017-5753 et CVE-2017-5754). CVE Pour plus d'informations, consultez <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versions des composants 5.11.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
aws-sagemaker-spark-sdk	1.0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.8.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.20.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème

Composant	Version	Description
		ème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-6	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-6	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-6	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-6	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.2-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.0.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	0.12.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.0.176	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.187	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.

Composant	Version	Description
presto-worker	0.187	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.1	Clients de ligne de commande Spark.
spark-history-server	2.2.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.11.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.11.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .

Classifications	Description
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue

Classifications	Description
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.

Classifications	Description
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.

Classifications	Description
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.11.0

Versions d'application 5.11.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
Kit AWS SDK pour Java	1,11,238	1,11,221	1,11,221	1,11,1183
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
HCatalog	2.3.2	2.3.1	2.3.1	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.2	2.3.1	2.3.1	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	0.12.0	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0
Presto	0.187	0.187	0.187	0.184
Spark	2.2.1	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.3	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.11.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.11.0 d'Amazon. Les modifications ont été apportées à la version 5.10.0.

Mises à niveau

- Hive 2.3.2
- Spark 2.2.1
- SDK pour Java 1.11.238

Nouvelles fonctionnalités

- Spark
 - Ajout du paramètre `spark.decommissioning.timeout.threshold`, qui améliore le comportement de mise hors service de Spark lors de l'utilisation d'instances Spot. Pour plus d'informations, consultez [Configuration du comportement de mise hors service du nœud](#).
 - Ajout du `aws-sagemaker-spark-sdk` composant à Spark, qui installe Amazon SageMaker Spark et les dépendances associées pour l'intégration de Spark à [Amazon SageMaker](#). Vous pouvez utiliser Amazon SageMaker Spark pour créer des pipelines d'apprentissage automatique (ML) Spark à l'aide d'Amazon SageMaker Stages. Pour plus d'informations, consultez le [fichier Readme de SageMaker Spark](#) sur GitHub et son [utilisation d'Apache Spark avec Amazon SageMaker](#) dans le manuel Amazon SageMaker Developer Guide.

Problèmes connus

- MXNet n'inclut pas les bibliothèques OpenCV.

- Hive 2.3.2 définit `hive.compute.query.using.stats=true` par défaut. Cela entraîne des requêtes pour obtenir des données à partir de statistiques existantes plutôt que directement à partir des données, ce qui peut être déroutant. Par exemple, si vous avez une table avec `hive.compute.query.using.stats=true` et que vous téléchargez de nouveaux fichiers vers la table LOCATION, l'exécution d'une demande `SELECT COUNT(*)` sur la table renvoie le nombre des statistiques, plutôt que de récupérer les lignes ajoutées.

Pour contourner ce problème, utilisez la commande `ANALYZE TABLE` pour collecter de nouvelles statistiques, ou définissez `hive.compute.query.using.stats=false`. Pour en savoir plus, consultez [Statistiques dans Hive](#) dans la documentation Apache Hive.

Versions des composants 5.11.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>aws-sagemaker-spark-sdk</code>	1.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.8.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.20.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-6	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.7.3-amzn-6	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.7.3-amzn-6	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-6	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.2-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.2-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.2-amzn-0	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.3.2-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.2-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.0.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	0.12.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.0.176	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.187	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.187	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.1	Clients de ligne de commande Spark.
spark-history-server	2.2.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.

Composant	Version	Description
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.11.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.11.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.

Classifications	Description
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.

Classifications	Description
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.

Classifications	Description
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.

Classifications	Description
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Classifications	Description
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.10.1

Versions d'application 5.10.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
Kit AWS SDK pour Java	1,11,221	1,11,221	1,11,1183	1,11,1183
Python	2,7, 3,4	2,7, 3,4	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.2
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.1	2.3.1	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.1	2.3.1	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	4.0.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	0,4,0
MXNet	0.12.0	0.12.0	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0.17.0

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
Presto	0.187	0.187	0.184	0.184
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.3	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.10.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.10.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.7.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.20.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-5	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-5	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-5	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-5	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-5	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-5	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.0.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	0.12.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.

Composant	Version	Description
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.0.176	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.187	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.187	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.0	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.2.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.10.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.10.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase

Classifications	Description
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.

Classifications	Description
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.10.0

Versions d'application 5.10.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
Kit AWS SDK pour Java	1,11,221	1,11,1183	1,11,1183	1,1,1160
Python	2,7, 3,4	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.2	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.1	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.1	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	4.0.1	3.12.0
Iceberg	-	-	-	-

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	0,4,0	-
MXNet	0.12.0	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0.17.0	0,16,0
Presto	0.187	0.184	0.184	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.3	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.10.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.10.0 d'Amazon. Les modifications concernent la EMR version 5.9.0 d'Amazon.

Mises à niveau

- AWS SDK for Java 1,11,221
- Hive 2.3.1
- Presto 0.187

Nouvelles fonctionnalités

- Ajout de la prise en charge de l'authentification Kerberos. Pour plus d'informations, consultez [Utiliser l'authentification Kerberos](#) dans le guide de gestion Amazon EMR
- Ajout de la prise en charge IAM des rôles pour EMRFS. Pour plus d'informations, consultez [Configurer IAM les rôles pour les EMRFS demandes adressées à Amazon S3](#) dans le guide EMR de gestion Amazon
- Ajout de la prise en charge GPU des types d'instances basés sur P2 et P3. Pour plus d'informations, consultez les sections Instances [Amazon EC2 P2 et Instances Amazon EC2 P3](#). NVIDIAle pilote 384.81 et CUDA le pilote 9.0.176 sont installés par défaut sur ces types d'instances.
- Ajout de la prise en charge de [Apache MXNet](#).

Modifications, améliorations et problèmes résolus

- Presto
 - Ajout de la prise en charge de l'utilisation du catalogue de données AWS Glue comme métastore Hive par défaut. Pour plus d'informations, consultez la section [Utilisation de Presto avec le catalogue de données AWS Glue](#).
 - Ajout de la prise en charge des [fonctions géospatiales](#).
 - Ajout de la prise en charge du [déversement sur le disque](#) pour les jointures.
 - Ajout de la prise en charge du [connecteur Redshift](#).
- Spark
 - Rétroporté [SPARK-20640](#), ce qui rend le délai d'expiration du RPC et les nouvelles tentatives d'enregistrement aléatoire configurables à l'aide des propriétés et des propriétés. `spark.shuffle.registration.timeout` `spark.shuffle.registration.maxAttempts`

- Rétroporté [SPARK-21549](#), qui corrige une erreur survenue lors de l'écriture personnalisée dans des emplacements non OutputFormat localisés. HDFS
- Rétroportage de [Hadoop-13270](#)
- Les bibliothèques Numpy, Scipy et Matplotlib ont été supprimées de la base Amazon. EMR AMI Si vous avez besoin de ces bibliothèques pour votre application, elles sont disponibles dans le référentiel d'applications. Vous pouvez utiliser une action d'amorçage pour les installer sur tous les nœuds à l'aide de `yum install`.
- La EMR base Amazon AMI ne contient plus de RPM packages d'applications, de sorte que les RPM packages ne sont plus présents sur les nœuds du cluster. Custom AMIs et la EMR base Amazon AMI font désormais référence au référentiel de RPM packages dans Amazon S3.
- En raison de l'introduction de la facturation à la seconde sur AmazonEC2, le comportement de réduction par défaut est désormais de terminer la tâche à la fin de la tâche plutôt que de terminer à l'heure de l'instance. Pour plus d'informations, consultez [Configuration de la diminution de la capacité des clusters](#).

Problèmes connus

- MXNet n'inclut pas les bibliothèques OpenCV.
- Hive 2.3.1 définit `hive.compute.query.using.stats=true` par défaut. Cela entraîne des requêtes pour obtenir des données à partir de statistiques existantes plutôt que directement à partir des données, ce qui peut être déroutant. Par exemple, si vous avez une table avec `hive.compute.query.using.stats=true` et que vous téléchargez de nouveaux fichiers vers la table LOCATION, l'exécution d'une demande `SELECT COUNT(*)` sur la table renvoie le nombre des statistiques, plutôt que de récupérer les lignes ajoutées.

Pour contourner ce problème, utilisez la commande `ANALYZE TABLE` pour collecter de nouvelles statistiques, ou définissez `hive.compute.query.using.stats=false`. Pour en savoir plus, consultez [Statistiques dans Hive](#) dans la documentation Apache Hive.

Versions des composants 5.10.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.5.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.7.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.20.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-5	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-5	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-5	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-5	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-5	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-5	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.0.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mxnet	0.12.0	Bibliothèque flexible, évolutive et efficace pour le deep learning.

Composant	Version	Description
mysql-server	5.5.54+	Mon serveur SQL de base de données.
nvidia-cuda	9.0.176	Pilotes Nvidia et boîte à outils Cuda
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.187	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.187	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.0	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.2.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.3	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.10.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.10.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase

Classifications	Description
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.

Classifications	Description
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.9.1

Versions d'application 5.9.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
Kit AWS SDK pour Java	1,11,1183	1,11,1183	1,1,1160	1,1,1160
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.2	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	4.0.1	4.0.1	3.12.0	3.12.0
Iceberg	-	-	-	-

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	0,4,0	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0.17.0	0,16,0	0,16,0
Presto	0.184	0.184	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.9.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

⚠ Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.9.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.4.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.7.0	Application de copie distribuée optimisée pour Amazon S3.

Composant	Version	Description
emrfs	2.19.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.2	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.7.3-amzn-4	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-4	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	2.3.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.0-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	4.0.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.184	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.184	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.17.0	Client de ligne de commande Pig.
spark-client	2.2.0	Clients de ligne de commande Spark.
spark-history-server	2.2.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.9.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.9.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop

Classifications	Description
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.

Classifications	Description
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.

Classifications	Description
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.

Classifications	Description
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.9.0

Versions d'application 5.9.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
Kit AWS SDK pour Java	1,11,1183	1,1,1160	1,1,1160	1,1,1160
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.2	1.3.1	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
Hue	4.0.1	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	0,4,0	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0.17.0	0,16,0	0,16,0	0,16,0
Presto	0.184	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.9.0

Les notes de publication suivantes contiennent des informations relatives à la EMR version 5.9.0 d'Amazon. Les modifications concernent la version EMR 5.8.0 d'Amazon.

Date de parution : 5 octobre 2017

Dernière date de mise à jour de la fonctionnalité : 12 octobre 2017

Mises à niveau

- AWS SDK for Java version 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

Nouvelles fonctionnalités

- Ajout de la prise en charge de Livy (version 0.4.0-incubating). Pour plus d'informations, consultez [Apache Livy](#).
- Ajout de la prise en charge de Hue Notebook pour Spark.
- Ajout de la prise en charge des EC2 instances Amazon de la série i3 (12 octobre 2017).

Modifications, améliorations et problèmes résolus

- Spark
 - Ajout d'un nouvel ensemble de fonctionnalités qui permet de s'assurer que Spark gère plus élégamment la terminaison des nœuds suite à un redimensionnement manuel ou à une demande de stratégie de dimensionnement automatique. Pour plus d'informations, consultez [Configuration du comportement de mise hors service du nœud](#).
 - SSL est utilisé au lieu de 3 DES pour le chiffrement en transit du service de transfert par blocs, ce qui améliore les performances lors de l'utilisation des types d'EC2 instances Amazon avec AES - NI.
 - Rétroporté [SPARK-21494](#).
- Zeppelin

- Rétroporté [ZEPPELIN-2377](#).
- HBase
 - Ajout du correctif [HBASE-18533](#), qui autorise des valeurs supplémentaires pour la configuration à l'aide HBase BucketCache de la classification de hbase-site configuration.
- Hue
 - Ajout de la prise en charge du catalogue de données AWS Glue pour l'éditeur de requêtes Hive dans Hue.
 - Par défaut, les superutilisateurs de Hue peuvent accéder à tous les fichiers auxquels les EMR IAM rôles Amazon sont autorisés à accéder. Les utilisateurs nouvellement créés n'ont pas automatiquement les autorisations d'accéder à l'explorateur de fichiers Amazon S3 et doivent avoir les autorisations `filebrowser.s3_access` activées pour leur groupe.
- Résolution d'un problème en raison duquel les JSON données sous-jacentes créées à l'aide de AWS Glue Data Catalog étaient inaccessibles.

Problèmes connus

- Le lancement du cluster échoue lorsque toutes les applications sont installées et que la taille du volume EBS racine Amazon par défaut n'est pas modifiée. Pour contourner le problème, utilisez la `aws emr create-cluster` commande du AWS CLI et spécifiez un `--ebs-root-volume-size` paramètre plus grand.
- Hive 2.3.0 définit `hive.compute.query.using.stats=true` par défaut. Cela entraîne des requêtes pour obtenir des données à partir de statistiques existantes plutôt que directement à partir des données, ce qui peut être déroutant. Par exemple, si vous avez une table avec `hive.compute.query.using.stats=true` et que vous téléchargez de nouveaux fichiers vers la table LOCATION, l'exécution d'une demande `SELECT COUNT(*)` sur la table renvoie le nombre des statistiques, plutôt que de récupérer les lignes ajoutées.

Pour contourner ce problème, utilisez la commande `ANALYZE TABLE` pour collecter de nouvelles statistiques, ou définissez `hive.compute.query.using.stats=false`. Pour en savoir plus, consultez [Statistiques dans Hive](#) dans la documentation Apache Hive.

Versions des composants 5.9.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à

Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.4.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.7.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.19.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.3.2	Applications et scripts client de la ligne de commande Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop.

Composant	Version	Description
		ème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.3-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.3.0-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	4.0.1	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
livy-server	0.4.0-incubating	RESTinterface pour interagir avec Apache Spark
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.184	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.184	Service pour exécuter les éléments d'une requête.
pig-client	0.17.0	Client de ligne de commande Pig.

Composant	Version	Description
spark-client	2.2.0	Clients de ligne de commande Spark.
spark-history-server	2.2.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.9.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.9.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase

Classifications	Description
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
livy-conf	Modifiez les valeurs dans le fichier livy.conf de Livy.
livy-env	Modifiez les valeurs dans l'environnement Livy.
livy-log4j	Modifiez les paramètres Livy log4j.properties.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-env	Modifiez les valeurs dans le fichier <code>presto-env.sh</code> de Presto.
presto-node	Modifiez les valeurs dans le fichier <code>node.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.

Classifications	Description
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.8.3

Versions d'application 5.8.3

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
Kit AWS SDK pour Java	1,1,1160	1,1,1160	1,1,1160	1,1,1160
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.3.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.3.0
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.8.3

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

⚠ Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.8.3

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.4.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.6.0	Application de copie distribuée optimisée pour Amazon S3.

Composant	Version	Description
emrfs	2.18.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.1	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.7.3-amzn-3	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.0-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.

Composant	Version	Description
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-1	Client de ligne de commande Pig.

Composant	Version	Description
spark-client	2.2.0	Clients de ligne de commande Spark.
spark-history-server	2.2.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.8.3

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.8.3

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase

Classifications	Description
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.8.2

Versions d'application 5.8.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
Kit AWS SDK pour Java	1,1,1160	1,1,1160	1,1,1160	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.1	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.3.0	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.3.0	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.2.0	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.8.2

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.8.2 d'Amazon. Les modifications ont été apportées à la version 5.8.1.

Date de parution initiale : 29 mars 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités potentielles.

Versions des composants 5.8.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.4.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.6.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.18.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.3.1	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.3-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.0-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-1	Client de ligne de commande Pig.
spark-client	2.2.0	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.2.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	2.2.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.8.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.8.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase

Classifications	Description
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.8.1

Versions d'application 5.8.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
Kit AWS SDK pour Java	1,1,1160	1,1,1160	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.1	1.3.0	1.3.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.1
HCatalog	2.3.0	2.3.0	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.3.0	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.11.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.2.0	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.8.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.8.1 d'Amazon. Les modifications concernent la version EMR 5.8.0 d'Amazon.

Date de parution initiale : 22 janvier 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités associées à l'exécution spéculative (CVE-2017-5715, -2017-5753 et CVE-2017-5754). CVE Pour plus d'informations, consultez <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versions des composants 5.8.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.4.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.6.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.18.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.3.1	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.3-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.3.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.3.0-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-1	Client de ligne de commande Pig.
spark-client	2.2.0	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.2.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.2.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.8.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.8.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase

Classifications	Description
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.

Classifications	Description
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.8.0

Versions d'application 5.8.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
Kit AWS SDK pour Java	1,1,1160	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.1	1.3.0	1.3.0	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.1	1.3.0
HCatalog	2.3.0	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.3.0	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.11.0	4.9.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.2.0	2.1.1	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.2	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.8.0

Les notes de publication suivantes contiennent des informations relatives à la EMR version 5.8.0 d'Amazon. Les modifications concernent la EMR version 5.7.0 d'Amazon.

Date de parution initiale : 10 août 2017

Dernière date de mise à jour de la fonctionnalité : 25 septembre 2017

Mises à niveau

- AWS SDK1,1,1160
- Flink 1.3.1
- Hive 2.3.0. Pour plus d'informations, consultez [Notes de mise à jour](#) sur le site Apache Hive.

- Spark 2.2.0. Pour plus d'informations, consultez [Notes de mise à jour](#) sur le site Apache Spark.

Nouvelles fonctionnalités

- Ajout de la prise en charge de l'affichage de l'historique de l'application (25 septembre 2017). Pour plus d'informations, consultez la section [Affichage de l'historique des applications](#) dans le guide EMR de gestion Amazon.

Modifications, améliorations et problèmes résolus

- Intégration avec AWS Glue Data Catalog
 - Hive et Spark ont désormais la possibilité SQL d'utiliser AWS Glue Data Catalog comme magasin de métadonnées Hive. Pour plus d'informations, consultez [Utiliser le catalogue de données AWS Glue comme métastore pour Hive](#) et [Utiliser le catalogue de données AWS Glue comme métastore pour Spark SQL](#).
- Ajout de l'historique des applications aux détails du cluster, ce qui vous permet de consulter les données historiques des YARN applications et des détails supplémentaires pour les applications Spark. Pour plus d'informations, consultez la section [Afficher l'historique des applications](#) dans le guide EMR de gestion Amazon.
- Oozie
 - Rétroporté [OOZIE-2748](#).
- Hue
 - [Rétroporté -5859 HUE](#)
- HBase
 - Ajout d'un correctif pour exposer l'heure de démarrage du serveur HBase principal via Java Management Extensions (JMX) à l'aide de `getMasterInitializedTime`.
 - Ajout d'un correctif qui améliore la date de début du cluster.

Problèmes connus

- Le lancement du cluster échoue lorsque toutes les applications sont installées et que la taille du volume EBS racine Amazon par défaut n'est pas modifiée. Pour contourner le problème, utilisez la `aws emr create-cluster` commande du AWS CLI et spécifiez un `--ebs-root-volume-size` paramètre plus grand.

- Hive 2.3.0 définit `hive.compute.query.using.stats=true` par défaut. Cela entraîne des requêtes pour obtenir des données à partir de statistiques existantes plutôt que directement à partir des données, ce qui peut être déroutant. Par exemple, si vous avez une table avec `hive.compute.query.using.stats=true` et que vous téléchargez de nouveaux fichiers vers la table `LOCATION`, l'exécution d'une demande `SELECT COUNT(*)` sur la table renvoie le nombre des statistiques, plutôt que de récupérer les lignes ajoutées.

Pour contourner ce problème, utilisez la commande `ANALYZE TABLE` pour collecter de nouvelles statistiques, ou définissez `hive.compute.query.using.stats=false`. Pour en savoir plus, consultez [Statistiques dans Hive](#) dans la documentation Apache Hive.

- Spark – Il existe un problème de fuite du gestionnaire de fichiers avec le démon `appusher` lors de l'utilisation de Spark, qui peut apparaître après plusieurs heures ou jours pour une tâche Spark de longue durée. Pour résoudre le problème, connectez-vous au nœud primaire et tapez `sudo /etc/init.d/appusher stop`. Cela arrête le daemon `appusher`, qu'Amazon EMR redémarrera automatiquement.
- Historique de l'application
 - Les données d'historique pour les exécuteurs Spark inactifs ne sont pas disponibles.
 - L'historique de l'application n'est pas disponible pour les clusters qui utilisent une configuration de sécurité pour permettre le chiffrement à la volée.

Versions des composants 5.8.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	4.4.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.4.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.4.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.6.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.18.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.1	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par

Composant	Version	Description
		le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.3.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.3.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.3.0-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.3.0-amzn-0	Client Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.3.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-1	Client de ligne de commande Pig.
spark-client	2.2.0	Clients de ligne de commande Spark.
spark-history-server	2.2.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.2.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.2.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.8.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.8.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog <code>proto-hive-site</code> fichier <code>.xml</code> .
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat <code>log4j2.properties</code> de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier <code>webhcat-site.xml</code> de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier <code>beeline-log4j2.properties</code> de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier <code>parquet-logging.properties</code> de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.

Classifications	Description
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.7.1

Versions d'application 5.7.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.3.0	1.2.1	1.2.1
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.1	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.13.0
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.11.0	4.9.0	4.9.0

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.1	2.1.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.2	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.7.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.7.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent

généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.18.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.3.0	Applications et scripts client de la ligne de commande Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.1	Clients de ligne de commande Spark.
spark-history-server	2.1.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.1	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.1.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.7.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.7.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.

Classifications	Description
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.7.0

Versions d'application 5.7.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.3.0	1.2.1	1.2.1	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.1	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.13.0	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.11.0	4.9.0	4.9.0	4.9.0

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.1	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.2	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.7.0

Les notes de mise à jour suivantes incluent des informations relatives à la version EMR 5.7.0 d'Amazon. Les modifications concernent la EMR version 5.6.0 d'Amazon.

Date de parution : 13 juillet 2017

Mises à niveau

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

Nouvelles fonctionnalités

- Ajout de la possibilité de spécifier un customAmazon Linux AMI lors de la création d'un cluster. Pour plus d'informations, consultez la section [Utilisation d'une option personnalisée AMI](#).

Modifications, améliorations et problèmes résolus

- HBase
 - Ajout de la possibilité de configurer des clusters HBase en lecture et réplication. Consultez [Utilisation d'un cluster de réplica en lecture](#).
 - Correctifs de nombreux bogues et améliorations
- Presto – ajout de la possibilité de configurer `node.properties`.
- YARN- possibilité ajoutée de configurer `container-log4j.properties`
- Sqoop - rétroporté [SQOOP-2880](#), qui introduit un argument vous permettant de définir le répertoire temporaire Sqoop.

Versions des composants 5.7.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.5.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.18.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.3.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,11,0- -1,3 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,11,0- -1,3 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.1	Clients de ligne de commande Spark.
spark-history-server	2.1.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.1.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.7.2	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.7.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.7.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .

Classifications	Description
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue

Classifications	Description
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.

Classifications	Description
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.

Classifications	Description
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.

Classifications	Description
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.6.1

Versions d'application 5.6.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.1	1.2.1	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.13.0	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.1	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.6.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.6.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.5.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.17.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.2.1	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-ftpfs-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,9,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,9,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.1	Clients de ligne de commande Spark.
spark-history-server	2.1.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.1.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.7.1	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.6.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.6.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .

Classifications	Description
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue

Classifications	Description
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.

Classifications	Description
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-env	Modifiez les valeurs dans le fichier presto-env.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.

Classifications	Description
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.

Classifications	Description
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.6.0

Versions d'application 5.6.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.1	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.13.0	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.1	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.6.0

Les notes de mise à jour suivantes incluent des informations relatives à la EMR version 5.6.0 d'Amazon. Les modifications concernent la version EMR 5.5.0 d'Amazon.

Date de parution : 5 juin 2017

Mises à niveau

- Flink 1.2.1
- HBase1.3.1
- Mahout 0.13.0. Il s'agit de la première version de Mahout compatible avec Spark 2.x dans Amazon EMR version 5.0 et versions ultérieures.
- Spark 2.1.1

Modifications, améliorations et problèmes résolus

- Presto
 - Ajout de la possibilité d'SSLTLsactiver/sécuriser la communication entre les nœuds Presto en activant le chiffrement en transit à l'aide d'une configuration de sécurité. Pour plus d'informations, consultez [Chiffrement des données en transit](#).
 - Rétroportage de [Presto 7661](#), ce qui ajoute l'option VERBOSE à l'instruction EXPLAIN ANALYZE pour transmettre des statistiques de bas niveau plus détaillées sur un plan de la requête.

Versions des composants 5.6.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par emr ou aws. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-

composant avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.17.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.2.1	Applications et scripts client de la ligne de commande Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.1	HBaseclient en ligne de commande.
hbase-rest-server	1.3.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.13.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,9,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,9,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.1	Clients de ligne de commande Spark.
spark-history-server	2.1.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.1	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.1.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.6.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.6.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-env	Modifiez les valeurs dans le fichier presto-env.v.sh de Presto.
presto-node	Modifiez les valeurs dans le fichier node.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.

Classifications	Description
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.5.4

Versions d'application 5.5.4

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.0	2.1.0	2.1.0	2.1.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.5.4

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.5.4

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent

généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.16.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.2.0	Applications et scripts client de la ligne de commande Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.0	HBaseclient en ligne de commande.
hbase-rest-server	1.3.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,9,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,9,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.5.4

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.5.4

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.5.3

Versions d'application 5.5.3

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.12.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.170
Spark	2.1.0	2.1.0	2.1.0	2.1.0

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.1
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notes de mise à jour 5.5.3

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.5.3 d'Amazon. Les modifications ont été apportées à la version 5.5.2.

Date de parution initiale : 29 août 2018

Modifications, améliorations et problèmes résolus

- Cette version corrige une vulnérabilité de sécurité potentielle.

Versions des composants 5.5.3

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-`

composant avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.16.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.2.0	Applications et scripts client de la ligne de commande Apache Flink.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.0	HBaseclient en ligne de commande.
hbase-rest-server	1.3.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,9,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,9,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.5.3

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.5.3

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.5.2

Versions d'application 5.5.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.12.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.170	0.166
Spark	2.1.0	2.1.0	2.1.0	2.1.0

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.1	0.7.0
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.9

Notes de mise à jour 5.5.2

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.5.2 d'Amazon. Les modifications ont été apportées à la version 5.5.1.

Date de parution initiale : 29 mars 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités potentielles.

Versions des composants 5.5.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-`

composant avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.5.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.16.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.2.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.0	HBaseclient en ligne de commande.
hbase-rest-server	1.3.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,9,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,9,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.5.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.5.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.5.1

Versions d'application 5.5.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.2.0
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.3.0
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.12.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.9.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.170	0.166	0.166
Spark	2.1.0	2.1.0	2.1.0	2.1.0

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.1	0.7.0	0.7.0
ZooKeeper	3.4.10	3.4.10	3.4.9	3.4.9

Notes de mise à jour 5.5.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.5.1 d'Amazon. Les modifications concernent la version EMR 5.5.0 d'Amazon.

Date de parution initiale : 22 janvier 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités associées à l'exécution spéculative (CVE-2017-5715, -2017-5753 et CVE-2017-5754). CVE Pour plus d'informations, consultez <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versions des composants 5.5.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.5.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.16.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.2.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les

Composant	Version	Description
		métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.0	HBaseclient en ligne de commande.
hbase-rest-server	1.3.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,9,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,9,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.5.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.5.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.5.0

Versions d'application 5.5.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.2.0	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.3.0	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.12.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.9.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.170	0.166	0.166	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.1	0.7.0	0.7.0	0.6.2
ZooKeeper	3.4.10	3.4.9	3.4.9	3.4.9

Notes de mise à jour 5.5.0

Les notes de mise à jour suivantes incluent des informations relatives à la version EMR 5.5.0 d'Amazon. Les modifications concernent la version EMR 5.4.0 d'Amazon.

Date de parution : 26 avril 2017

Mises à niveau

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

Modifications, améliorations et problèmes résolus

- Spark
 - Patch Spark rétroporté ([SPARK-20115](#)) [Correction DAGScheduler permettant de recalculer tous les blocs de shuffle perdus lorsque le service de shuffle externe n'est pas disponible](#) dans la version 2.1.0 de Spark, incluse dans cette version.
- Flink
 - Flink s'appuie désormais sur Scala 2.11. Si vous utilisez Scala API et les bibliothèques, nous vous recommandons d'utiliser Scala 2.11 dans vos projets.

- Résolution d'un problème où les valeurs par défaut HADOOP_CONF_DIR et YARN_CONF_DIR n'étaient pas définies correctement, si bien que `start-scala-shell.sh` ne fonctionnait pas. Ajout également de la possibilité de définir ces valeurs à l'aide de `env.hadoop.conf.dir` et `env.yarn.conf.dir` dans `/etc/flink/conf/flink-conf.yaml` ou la classification de configuration `flink-conf`.
- Introduction d'une nouvelle commande EMR spécifique, `flink-scala-shell` en tant que wrapper pour `start-scala-shell.sh`. Nous vous recommandons d'utiliser cette commande au lieu de `start-scala-shell`. La nouvelle commande simplifie l'exécution. Par exemple, `flink-scala-shell -n 2` démarre un shell Flink Scala avec un parallélisme de 2 tâches.
- Introduction d'une nouvelle commande EMR spécifique, `flink-yarn-session` en tant que wrapper pour `yarn-session.sh`. Nous vous recommandons d'utiliser cette commande au lieu de `yarn-session`. La nouvelle commande simplifie l'exécution. Par exemple, `flink-yarn-session -d -n 2` démarre une session Flink de longue durée à l'état détaché avec deux gestionnaires de tâches.
- Adressé ([FLINK-6125](#)) [Le client HTTP de Commons n'est plus ombré dans](#) Flink 1.2.
- Presto
 - Ajout de la prise en charge de LDAP l'authentification. L'utilisation LDAP avec Presto sur Amazon EMR nécessite que vous autorisiez l'HTTPS accès au coordinateur Presto (`http-server.https.enabled=true` in `config.properties`). Pour les détails de configuration, voir [LDAPAuthentification](#) dans la documentation Presto.
 - Ajout de la prise en charge de `SHOW GRANTS`.
- Amazon EMR Base Linux AMI
 - Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2017.03. Pour plus d'informations, consultez les notes de [mise à jour d'Amazon Linux AMI 2017.03](#).
 - Python 2.6 a été supprimé de l'image Linux EMR de base d'Amazon. Python 2.7 et 3.4 sont installés par défaut. Vous pouvez installer Python 2.6 manuellement si nécessaire.

Versions des composants 5.5.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version

d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.5.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.16.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.2.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.0	HBaseclient en ligne de commande.
hbase-rest-server	1.3.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.12.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,9,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,9,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.170	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.170	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.10	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.10	ZooKeeper client en ligne de commande.

Classifications des configurations 5.5.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.5.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.4.1

Versions d'application 5.4.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
Delta	-	-	-	-
Flink	1.2.0	1.2.0	1.1.4	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.3.0	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.9.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.166	0.166	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.0	0.7.0	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

Notes de mise à jour 5.4.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.4.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion`

commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.15.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.2.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.3.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.0	HBaseclient en ligne de commande.
hbase-rest-server	1.3.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,9,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,9,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.166	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.166	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.7.0	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

Classifications des configurations 5.4.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.4.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.4.0

Versions d'application 5.4.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
Delta	-	-	-	-
Flink	1.2.0	1.1.4	1.1.4	1.1.4
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.3.0	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.1
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.1
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.3.0
Phoenix	4.9.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.166	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.1.0

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.7.0	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

Notes de mise à jour 5.4.0

Les notes de mise à jour suivantes incluent des informations relatives à la version EMR 5.4.0 d'Amazon. Les modifications concernent la version EMR 5.3.0 d'Amazon.

Date de parution : 8 mars 2017

Mises à niveau

- Mise à niveau vers Flink 1.2.0
- Mise à niveau vers la version HBase 1.3.0
- Mise à niveau vers Phoenix 4.9.0

Note

Si vous effectuez une mise à niveau d'une version antérieure d'Amazon EMR vers la EMR version 5.4.0 ou ultérieure d'Amazon et que vous utilisez l'indexation secondaire, mettez à niveau les index locaux comme décrit dans la documentation d'[Apache Phoenix](#). Amazon EMR supprime les configurations requises de la `hbase-site` classification, mais les index doivent être remplis. La mise à niveau en ligne et hors ligne des index est prise en charge. Les mises à niveau sont par défaut en ligne, ce qui signifie que les index sont remplis de nouveau lors de l'initialisation depuis les clients Phoenix en version 4.8.0 ou supérieure. Pour spécifier les mises à niveau hors ligne, définissez la

```
phoenix.client.localIndexUpgrade configuration sur false dans la phoenix-site
classification, puis SSH sur le nœud principal à exécuterpsql [zookeeper] -1.
```

- Mise à niveau vers Presto 0.166
- Mise à niveau vers Zeppelin 0.7.0

Modifications et améliorations

- Ajout de la prise en charge des instances r4. Consultez la section [Types d'EC2instances Amazon](#).

Versions des composants 5.4.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.15.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.2.0	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-ftpfs-server	2.7.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.3.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.3.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.3.0	HBaseclient en ligne de commande.
hbase-rest-server	1.3.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.3.0	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-hbase	2.1.1-amzn-0	Client Hive-hbase.

Composant	Version	Description
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server2	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,9,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,9,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API

Composant	Version	Description
presto-coordinator	0.166	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.166	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.7.0	Portable basé sur le web qui permet l'analyse de données interactives.

Composant	Version	Description
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

Classifications des configurations 5.4.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.4.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .

Classifications	Description
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue

Classifications	Description
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.

Classifications	Description
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.

Classifications	Description
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.

Classifications	Description
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.3.2

Versions d'application 5.3.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)

- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.4	1.1.4	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.1	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.1	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.11.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.3.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.1.0	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

Notes de mise à jour 5.3.2

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

⚠ Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.3.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Application de copie distribuée optimisée pour Amazon S3.

Composant	Version	Description
emrfs	2.14.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.1.4	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.7.3-amzn-1	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.3	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.3	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.2.3	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.2.3	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop.

Composant	Version	Description
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.6.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

Classifications des configurations 5.3.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.3.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase

Classifications	Description
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.

Classifications	Description
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.

Classifications	Description
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.

Classifications	Description
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.3.1

Versions d'application 5.3.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.4	1.1.4	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.1	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.1	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.11.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.3.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.1.0	2.0.2	2.0.2
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

Notes de mise à jour 5.3.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.3.1 d'Amazon. Les modifications concernent la version EMR 5.3.0 d'Amazon.

Date de version : 7 février 2017

Des modifications mineures ont été apportées au rétroportage des correctifs Zeppelin et à la mise à jour de la valeur par défaut pour AmazonAMI. EMR

Versions des composants 5.3.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.14.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.1.4	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les

Composant	Version	Description
		métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.3	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.3	HBaseclient en ligne de commande.
hbase-rest-server	1.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.3	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.1-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.6.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

Classifications des configurations 5.3.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.3.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.3.0

Versions d'application 5.3.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Delta	-	-	-	-
Flink	1.1.4	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.1	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.1	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.11.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.3.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.157.1
Spark	2.1.0	2.0.2	2.0.2	2.0.2

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

Notes de mise à jour 5.3.0

Les notes de mise à jour suivantes incluent des informations relatives à la version EMR 5.3.0 d'Amazon. Les modifications concernent la version EMR 5.2.1 d'Amazon.

Date de parution : 26 janvier 2017

Mises à niveau

- Mise à niveau vers Hive 2.1.1
- Mise à niveau vers Hue 3.11.0
- Mise à niveau vers Spark 2.1.0
- Mise à niveau vers Oozie 4.3.0
- Mise à niveau vers Flink 1.1.4

Modifications et améliorations

- Ajout à Hue d'un correctif qui vous permet d'utiliser le paramètre `interpreters_shown_on_wheel` pour configurer les interpréteurs à afficher en premier sur la roue de sélection Notebook quel que soit leur ordre dans le fichier `hue.ini`.
- Ajout de la classification de configuration, `hive-parquet-logging` que vous pouvez utiliser pour configurer des valeurs dans le fichier `parquet-logging.properties` de Hive.

Versions des composants 5.3.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.14.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.1.4	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.3	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.3	HBaseclient en ligne de commande.
hbase-rest-server	1.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.2.3	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.1.1-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.1.1-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.1-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.1-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.11.0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.

Composant	Version	Description
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.3.0	Client de ligne de commande Oozie.
oozie-server	4.3.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.1.0	Clients de ligne de commande Spark.
spark-history-server	2.1.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	2.1.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.1.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.6.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

Classifications des configurations 5.3.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.3.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-parquet-logging	Modifiez les valeurs dans le fichier parquet-logging.properties de Hive.

Classifications	Description
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.2.3

Versions d'application 5.2.3

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.157.1	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.2

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.8

Notes de mise à jour 5.2.3

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.2.3

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion`

commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.13.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.1.3	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.

Composant	Version	Description
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.3	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.3	HBaseclient en ligne de commande.
hbase-rest-server	1.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.3	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantiqu e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applica tions de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technolog ie Machine Learning.
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.2	Clients de ligne de commande Spark.
spark-history-server	2.0.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.0.2	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.0.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTP serveur Apache.
zeppelin-server	0.6.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

Classifications des configurations 5.2.3

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.2.3

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.

Classifications	Description
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.

Classifications	Description
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.

Classifications	Description
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.2.2

Versions d'application 5.2.2

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.157.1	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.2	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.8	3.4.8

Notes de mise à jour 5.2.2

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.2.2 d'Amazon. Les modifications concernent la version EMR 5.2.1 d'Amazon.

Date de parution : 2 mai 2017

Problèmes connus résolus depuis les versions précédentes

- Rétroporté [SPARK-194459](#), qui résout un problème en cas d'échec de la lecture d'une ORC table contenant des colonnes char/varchar.

Versions des composants 5.2.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.13.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.1.3	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par

Composant	Version	Description
		le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.3	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.3	HBaseclient en ligne de commande.
hbase-rest-server	1.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.3	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantiqu e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applica tions de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technolog ie Machine Learning.
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.2	Clients de ligne de commande Spark.
spark-history-server	2.0.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.0.2	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.0.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTP serveur Apache.
zeppelin-server	0.6.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

Classifications des configurations 5.2.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.2.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.

Classifications	Description
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.

Classifications	Description
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.

Classifications	Description
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.2.1

Versions d'application 5.2.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	1.1.3
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.3

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.157.1	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.2	2.0.1	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.8	3.4.8	3.4.8

Notes de mise à jour 5.2.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.2.1 d'Amazon. Les modifications concernent la version EMR 5.2.0 d'Amazon.

Date de parution : 29 décembre 2016

Mises à niveau

- Mise à niveau vers Presto 0.157.1. Pour plus d'informations, consultez [Notes de mise à jour Presto](#) dans la documentation Presto.
- Mise à niveau vers Zookeeper 3.4.9. Pour plus d'informations, consultez les [notes ZooKeeper de version](#) dans la ZooKeeper documentation d'Apache.

Modifications et améliorations

- Ajout de la prise en charge du type d'EC2instance Amazon m4.16xlarge dans les EMR versions 4.8.3 et ultérieures d'Amazon, à l'exception des versions 5.0.0, 5.0.3 et 5.2.0.
- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.09. Pour plus d'informations, consultez <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.
- L'emplacement de Flink et les chemins de YARN configuration sont désormais définis par défaut, `/etc/default/flink` ce qui signifie qu'il n'est pas nécessaire de définir les variables d'environnement `FLINK_CONF_DIR` et `HADOOP_CONF_DIR` lorsque vous exécutez les scripts du `yarn-session.sh` pilote `flink` ou pour lancer les tâches Flink.
- Ajout du support pour les FlinkKinesisConsumer cours.

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème dans Hadoop où le ReplicationMonitor thread pouvait rester bloqué pendant longtemps en raison d'une course entre la réplication et la suppression du même fichier dans un grand cluster.
- Correction d'un problème en raison duquel ControlledJob # toString échouait avec une exception de pointeur nul (NPE) lorsque le statut de la tâche n'était pas correctement mis à jour.

Versions des composants 5.2.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.13.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.1.3	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.7.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.2.3	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.3	HBaseclient en ligne de commande.
hbase-rest-server	1.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.3	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.2	Clients de ligne de commande Spark.
spark-history-server	2.0.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.0.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.0.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.6.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

Classifications des configurations 5.2.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.2.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop

Classifications	Description
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.

Classifications	Description
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.

Classifications	Description
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.

Classifications	Description
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.

Classifications	Description
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.2.0

Versions d'application 5.2.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	1.1.3	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.3	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.152.3	0.152.3	0.152.3
Spark	2.0.2	2.0.1	2.0.1	2.0.1
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

Notes de mise à jour 5.2.0

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.2.0 d'Amazon. Les modifications concernent la version EMR 5.1.0 d'Amazon.

Date de parution : 21 novembre 2016

Modifications et améliorations

- Ajout du mode de stockage Amazon S3 pour HBase.
- Vous permet de spécifier un emplacement Amazon S3 pour le HBase rootdir. Pour plus d'informations, consultez [HBaseAmazon S3](#).

Mises à niveau

- Mise à niveau vers Spark 2.0.2

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème lié à la limitation de /mnt à 2 To pour les types d'instance EBS réservés uniquement.
- Correction d'un problème avec les journaux instance-controller et logpusher édités dans leurs fichiers .out correspondants au lieu de leurs fichiers .log configurés log4j, qui permutent toutes les heures. Les fichiers .out n'effectuent pas de permutation, cela remplirait finalement la partition /emr. Ce problème concerne uniquement les types d'instances matérielles de machine virtuelle (HVM).

Versions des composants 5.2.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.12.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.1.3	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.7.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.2.3	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.3	HBaseclient en ligne de commande.
hbase-rest-server	1.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.3	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.152.3	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.152.3	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.2	Clients de ligne de commande Spark.
spark-history-server	2.0.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.0.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.0.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.6.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

Classifications des configurations 5.2.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.2.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres <code>flink-conf.yaml</code> .
flink-log4j	Modifiez les paramètres <code>log4j.properties</code> Flink.
flink-log4j-yarn-session	Modifiez les paramètres du <code>j-yarn-session</code> fichier <code>.properties</code> de Flink log4.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink <code>log4j-cli</code> .
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop

Classifications	Description
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase	Paramètres EMR sélectionnés par Amazon pour Apache. HBase
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.

Classifications	Description
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.

Classifications	Description
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.

Classifications	Description
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.

Classifications	Description
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.1.1

Versions d'application 5.1.1

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	1.1.3	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.3	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.152.3	0.152.3	0.150
Spark	2.0.1	2.0.1	2.0.1	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

Notes de mise à jour 5.1.1

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 5.1.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.11.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>flink-client</code>	1.1.3	Applications et scripts client de la ligne de commande Apache Flink.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.3	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.3	HBaseclient en ligne de commande.
hbase-rest-server	1.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.2.3	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.

Composant	Version	Description
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.152.3	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.152.3	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.1	Clients de ligne de commande Spark.
spark-history-server	2.0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	2.0.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTP serveur Apache.
zeppelin-server	0.6.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

Classifications des configurations 5.1.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.1.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.

Classifications	Description
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.

Classifications	Description
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.

Classifications	Description
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.

Classifications	Description
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.1.0

Versions d'application 5.1.0

Les applications suivantes sont prises en charge dans cette version : [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	1.1.3	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.3	1.2.2	1.2.2	1.2.2

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.3	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.152.3	0.150	0.150
Spark	2.0.1	2.0.1	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

Notes de mise à jour 5.1.0

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.1.0 d'Amazon. Les modifications concernent la version EMR 5.0.3 d'Amazon.

Date de parution : 3 novembre 2016

Modifications et améliorations

- Support ajouté pour Flink 1.1.3.
- Presto a été ajouté comme une option dans la section ordinateur portable de Hue.

Mises à niveau

- Mise à niveau vers la version HBase 1.2.3
- Mise à niveau vers Zeppelin 0.6.2

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème lié aux requêtes Tez sur Amazon S3 avec des ORC fichiers qui ne fonctionnaient pas aussi bien que dans les versions antérieures d'Amazon EMR 4.x.

Versions des composants 5.1.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus

récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.11.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
flink-client	1.1.3	Applications et scripts client de la ligne de commande Apache Flink.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.

Composant	Version	Description
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.3	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.3	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.2.3	HBaseclient en ligne de commande.
hbase-rest-server	1.2.3	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.3	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	2.1.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantiqu e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applica tions de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technolog ie Machine Learning.
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.152.3	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.152.3	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.1	Clients de ligne de commande Spark.
spark-history-server	2.0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.0.1	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTP serveur Apache.
zeppelin-server	0.6.2	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

Classifications des configurations 5.1.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.1.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
flink-conf	Modifiez les paramètres flink-conf.yaml.
flink-log4j	Modifiez les paramètres log4j.properties Flink.
flink-log4j-yarn-session	Modifiez les paramètres du j-yarn-session fichier .properties de Flink log4j.
flink-log4j-cli	Modifiez les paramètres de propriétés Flink log4j-cli.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.

Classifications	Description
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.

Classifications	Description
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.

Classifications	Description
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.

Classifications	Description
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.0.3

Versions d'application 5.0.3

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

Notes de mise à jour 5.0.3

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.0.3 d'Amazon. Les modifications concernent la version EMR 5.0.0 d'Amazon.

Date de parution : 24 octobre 2016

Mises à niveau

- Mise à niveau vers Hadoop 2.7.3
- Mise à niveau vers Presto 0.152.3, qui comprend la prise en charge de l'interface Web Presto. Utilisez le port 8889 du coordinateur Presto pour accéder à l'interface Web Presto. Pour plus d'informations sur l'interface Web Presto, consultez [Interface Web](#) dans la documentation Presto.
- Mise à niveau vers Spark 2.0.1
- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.09. Pour plus d'informations, consultez <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Versions des composants 5.0.3

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-`

composant avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.10.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.152.3	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.152.3	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.1	Clients de ligne de commande Spark.
spark-history-server	2.0.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.0.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.0.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.

Composant	Version	Description
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

Classifications des configurations 5.0.3

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.0.3

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.

Classifications	Description
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.

Classifications	Description
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.

Classifications	Description
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.

Classifications	Description
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.0.2

Versions d'application 5.0.2

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

Notes de mise à jour 5.0.2

Versions des composants 5.0.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	4.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.9.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.2-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.2-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.46	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.150	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.150	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.0	Clients de ligne de commande Spark.
spark-history-server	2.0.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.0.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.0.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.

Composant	Version	Description
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0,6,1- SNAPSHOT	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

Classifications des configurations 5.0.2

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.0.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.

Classifications	Description
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.

Classifications	Description
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.

Classifications	Description
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.

Classifications	Description
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.0.1

Versions d'application 5.0.1

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

Notes de mise à jour 5.0.1

Versions des composants 5.0.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	4.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.9.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.2-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.2-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.46	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.150	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.150	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.0	Clients de ligne de commande Spark.
spark-history-server	2.0.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.0.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	2.0.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.

Composant	Version	Description
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0,6,1- SNAPSHOT	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

Classifications des configurations 5.0.1

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.0.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log 4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log 4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.

Classifications	Description
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.

Classifications	Description
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.

Classifications	Description
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.

Classifications	Description
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 5.0.0

Versions d'application 5.0.0

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) et [ZooKeeper](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	2.1.0	2.1.0	2.1.0	2.1.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	2.1.0	2.1.0	2.1.0	2.1.0
Hudi	-	-	-	-
Hue	3.10.0	3.10.0	3.10.0	3.10.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	4.2.0	4.2.0	4.2.0	4.2.0

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,16,0	0,16,0	0,16,0	0,16,0
Presto	0.152.3	0.150	0.150	0.150
Spark	2.0.1	2.0.0	2.0.0	2.0.0
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	3.4.8	3.4.8	3.4.8	3.4.8

Notes de mise à jour 5.0.0

Date de parution : 27 juillet 2016

Mises à niveau

- Mise à niveau vers Hive 2.1
- Mise à niveau vers Presto 0.150
- Mise à niveau vers Spark 2.0
- Mise à niveau vers Hue 3.10.0
- Mise à niveau vers Pig 0.16.0
- Mise à niveau vers Tez 0.8.4
- Mise à niveau vers Zeppelin 0.6.1

Modifications et améliorations

- Amazon EMR prend en charge les dernières versions open source de Hive (version 2.1) et Pig (version 0.16.0). Si vous avez déjà utilisé Hive ou Pig sur AmazonEMR, cela peut affecter certains cas d'utilisation. Pour en savoir plus, consultez [Hive](#) et [Pig](#).
- Tez est désormais le moteur d'exécution par défaut de Hive et Pig. Pour changer cela, vous devez modifier les valeurs appropriées dans les classifications de configuration `hive-site` et `pig-properties`, respectivement.
- Une fonction améliorée de débogage des étapes a été ajoutée, ce qui vous permet de voir la cause première des échecs des étapes si le service peut déterminer la cause. Pour plus d'informations, consultez [Enhanced Step Debugging](#) dans le Amazon EMR Management Guide.
- Les applications qui se terminaient par « -Sandbox » n'ont plus ce suffixe. Cela peut interrompre l'automatisation de certaines tâches, par exemple, si vous utilisez des scripts pour lancer des clusters avec ces applications. Le tableau suivant montre les noms des applications dans Amazon EMR 4.7.2 par rapport à Amazon EMR 5.0.0.

Modification des noms d'application

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Bac à sable	ZooKeeper

- Spark est maintenant compilé pour Scala 2.11.
- Java 8 est désormais la version par défaut JVM. Toutes les applications s'exécutent avec le runtime Java 8. Aucune modification n'est apportée bytecode cible des applications. La plupart des applications continuent de cibler Java 7.
- Zeppelin inclut désormais des fonctions d'authentification. Pour en savoir plus, consultez [Zeppelin](#).
- Ajout de la prise en charge des configurations de sécurité, qui vous permettent de créer et d'appliquer des options de chiffrement plus facilement. Pour en savoir plus, consultez [Chiffrement des données](#).

Versions des composants 5.0.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la EMR version la plus récente d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.9.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème

Composant	Version	Description
		ème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.2-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.2-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service pour desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour. HBase

Composant	Version	Description
hcatalog-client	2.1.0-amzn-0	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	2.1.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	2.1.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	2.1.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.10.0-amzn-0	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.46	Mon serveur SQL de base de données.

Composant	Version	Description
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.150	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.150	Service pour exécuter les éléments d'une requête.
pig-client	0.16.0-amzn-0	Client de ligne de commande Pig.
spark-client	2.0.0	Clients de ligne de commande Spark.
spark-history-server	2.0.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	2.0.0	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	2.0.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTP serveur Apache.
zeppelin-server	0,6,1- SNAPSHOT	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

Classifications des configurations 5.0.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-5.0.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.

Classifications	Description
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j2	Modifiez les valeurs dans le fichier ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-beeline-log4j2	Modifiez les valeurs dans le fichier beeline-log4j2.properties de Hive.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j2	Modifiez les valeurs dans le fichier hive-exec-log4j2.properties de Hive.
hive-llap-daemon-log4j2	Modifiez les valeurs dans le fichier llap-daemon-log4j2.properties de Hive.
hive-log4j2	Modifiez les valeurs dans le fichier hive-log4j2.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2

Classifications	Description
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.

Classifications	Description
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.

Classifications	Description
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Spark
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Versions publiées par Amazon EMR 4.x

Cette section contient les versions des applications, les notes de mise à jour, les versions des composants et les classifications de configuration disponibles dans chaque version d'Amazon EMR 4.x.

Lorsque vous lancez un cluster, vous pouvez choisir parmi plusieurs versions d'AmazonEMR. Vous pouvez ainsi tester et utiliser des versions d'applications qui répondent à vos exigences de compatibilité. Vous spécifiez le numéro de version avec l'étiquette de version. Les étiquettes de version sont sous la forme `emr-x.x.x`. Par exemple, `emr-7.2.0`.

Les nouvelles EMR versions d'Amazon sont mises à disposition dans différentes régions sur une période de plusieurs jours, en commençant par la première région à la date de sortie initiale. Il est possible que la dernière version ne soit pas disponible dans votre région pendant cette période.

Pour un tableau complet des versions des applications dans chaque version EMR 4.x d'Amazon, consultez [Versions des applications dans les versions 4.x d'Amazon EMR](#).

Rubriques

- [Versions des applications dans les versions 4.x d'Amazon EMR](#)
- [Différences entre les versions Amazon EMR 4.x](#)
- [Amazon EMR version 4.9.6](#)
- [Amazon EMR version 4.9.5](#)
- [Amazon EMR version 4.9.4](#)
- [Amazon EMR version 4.9.3](#)
- [Amazon EMR version 4.9.2](#)
- [Amazon EMR version 4.9.1](#)
- [Amazon EMR version 4.8.5](#)
- [Amazon EMR version 4.8.4](#)
- [Amazon EMR version 4.8.3](#)
- [Amazon EMR version 4.8.2](#)
- [Amazon EMR version 4.8.1](#)
- [Amazon EMR version 4.8.0](#)
- [Amazon EMR version 4.7.4](#)
- [Amazon EMR version 4.7.3](#)
- [Amazon EMR version 4.7.2](#)
- [Amazon EMR version 4.7.1](#)
- [Amazon EMR version 4.7.0](#)
- [Amazon EMR version 4.6.1](#)
- [Amazon EMR version 4.6.0](#)
- [Amazon EMR version 4.5.0](#)
- [Amazon EMR version 4.4.0](#)
- [Amazon EMR version 4.3.0](#)

- [Amazon EMR version 4.2.0](#)
- [Amazon EMR version 4.1.0](#)
- [Amazon EMR version 4.0.0](#)

Versions des applications dans les versions 4.x d'Amazon EMR

Pour obtenir un tableau complet répertoriant les versions d'applications disponibles dans chaque version d'Amazon EMR 4.x, ouvrez [Versions des applications dans les versions 4.x d'Amazon EMR](#) dans votre navigateur.

Différences entre les versions Amazon EMR 4.x

La documentation relative aux fonctionnalités d'Amazon EMR dans le Guide de gestion Amazon EMR précise la version d'Amazon EMR dans laquelle une fonctionnalité est devenue disponible, ainsi que les différences applicables entre les fonctionnalités d'Amazon EMR qui remontent à la version 4.0.0.

À partir de la version Amazon EMR 5.0.0, certaines applications ont subi une mise à niveau de version importante qui a modifié les modalités d'installation ou opérationnelles, et d'autres applications sont passées du statut d'application sandbox au statut d'application native. Chaque rubrique de cette section présente les différences significatives spécifiques aux applications lorsque vous utilisez les versions Amazon EMR 4.x.

Rubriques

- [Applications de l'environnement de test \(sandbox\)](#)
- [Considérations relatives à l'utilisation de Hive sur Amazon EMR 4.x](#)
- [Considérations relatives à l'utilisation de Pig sur Amazon EMR 4.x](#)

Applications de l'environnement de test (sandbox)

Lorsque vous utilisez les versions 4.x d'Amazon EMR, certaines applications sont considérées comme des applications d'environnement de test (sandbox). Les applications de l'environnement de test (sandbox) sont des versions antérieures de l'application que nous avons mises à disposition au moment du lancement de la première version d'Amazon EMR en raison de la demande. Vous pouvez utiliser la console, l'AWS CLI ou l'API pour qu'Amazon EMR installe les applications d'environnement de test (sandbox) de la même manière que les applications natives, mais la prise en charge et la documentation pour les applications d'environnement de test (sandbox) sont limitées. Les

applications sandbox deviennent des applications natives et entièrement prises en charge dans la version Amazon EMR 5.0.0 et versions ultérieures. Voici des applications sandbox dans les versions Amazon EMR 4.x :

- Oozie
- Presto
- Sqoop
- Zeppelin
- ZooKeeper

Lorsque vous installez des applications sandbox, le nom des applications porte le suffixe -sandbox. Par exemple, pour installer la version d'environnement de test (sandbox) de *Presto*, utilisez le nom `Presto-sandbox`. L'installation de cette application peut prendre plus de temps que celle d'une application entièrement prise en charge. Les numéros de version répertoriés pour chaque application dans cette section correspondent à la version de communauté de l'application.

Oozie (versions de l'environnement de test)

Oozie est disponible sous la forme d'une application d'environnement de test (sandbox) à partir de la version Amazon EMR 4.1.0.

Les exemples Oozie ne sont pas installés par défaut avec les versions sandbox. Pour installer les exemples, accédez au nœud principal via SSH et exécutez `install-oozie-examples`.

Informations sur la version Oozie-Sandbox

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.9.6	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resour

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
		cemanager, oozie-client, oozie-server
emr-4.9.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, oozie-client, oozie-server
emr-4.9.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, oozie-client, oozie-server

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.9.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.9.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.9.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.5	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.8.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server
emr-4.8.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, oozie-client, oozie-server

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.8.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.8.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.8.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.7.4	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.7.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.7.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.7.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.7.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.6.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.6.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.5.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.4.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Étiquette de version Amazon EMR	Oozie – Version d'environnement de test (sandbox)	Composants installés avec Oozie-Sandbox
emr-4.3.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.2.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server
emr-4.1.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Presto (versions de l'environnement de test)

Presto est disponible sous la forme d'une application d'environnement de test (sandbox) à partir de la version Amazon EMR 4.1.0.

Informations sur la version Presto-Sandbox

Étiquette de version Amazon EMR	Presto – Version d'environnement de test (sandbox)	Composants installés avec Presto-Sandbox
emr-4.9.6	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.5	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.4	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client,

Étiquette de version Amazon EMR	Presto – Version d'environnement de test (sandbox)	Composants installés avec Presto-Sandbox
		hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.9.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette de version Amazon EMR	Presto – Version d'environnement de test (sandbox)	Composants installés avec Presto-Sandbox
emr-4.9.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.5	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.4	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette de version Amazon EMR	Presto – Version d'environnement de test (sandbox)	Composants installés avec Presto-Sandbox
emr-4.8.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.2	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.8.1	0.151	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette de version Amazon EMR	Presto – Version d'environnement de test (sandbox)	Composants installés avec Presto-Sandbox
emr-4.8.0	0.151	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.4	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.3	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette de version Amazon EMR	Presto – Version d'environnement de test (sandbox)	Composants installés avec Presto-Sandbox
emr-4.7.2	0.148	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.1	0.147	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-4.7.0	0.147	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette de version Amazon EMR	Presto – Version d'environnement de test (sandbox)	Composants installés avec Presto-Sandbox
emr-4.6.1	0.143	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.6.0	0.143	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.5.0	0.140	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Étiquette de version Amazon EMR	Presto – Version d'environnement de test (sandbox)	Composants installés avec Presto-Sandbox
emr-4.4.0	0.136	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.3.0	0.130	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker
emr-4.2.0	0.125	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Étiquette de version Amazon EMR	Presto – Version d'environnement de test (sandbox)	Composants installés avec Presto-Sandbox
emr-4.1.0	0.119	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, mysql-server, presto-coordinator, presto-worker

Sqoop (versions de l'environnement de test)

Sqoop est disponible sous la forme d'une application d'environnement de test (sandbox) à partir de la version Amazon EMR 4.4.0.

Informations sur la version Sqoop-Sandbox

Étiquette de version Amazon EMR	Sqoop – Version d'environnement de test (sandbox)	Composants installés avec Sqoop-Sandbox
emr-4.9.6	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-

Étiquette de version Amazon EMR	Sqoop – Version d'environnement de test (sandbox)	Composants installés avec Sqoop-Sandbox
		hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema- nager, hadoop-yarn-resour- cemanager, mysql-server, sqoop-client
emr-4.9.4	1.4.6	emrfs, emr-ddb, emr-goodi- es, hadoop-client, hadoop-ma- pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema- nager, hadoop-yarn-resour- cemanager, mysql-server, sqoop-client
emr-4.9.3	1.4.6	emrfs, emr-ddb, emr-goodi- es, hadoop-client, hadoop-ma- pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema- nager, hadoop-yarn-resour- cemanager, mysql-server, sqoop-client

Étiquette de version Amazon EMR	Sqoop – Version d'environnement de test (sandbox)	Composants installés avec Sqoop-Sandbox
emr-4.9.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.5	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette de version Amazon EMR	Sqoop – Version d'environnement de test (sandbox)	Composants installés avec Sqoop-Sandbox
emr-4.8.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette de version Amazon EMR	Sqoop – Version d'environnement de test (sandbox)	Composants installés avec Sqoop-Sandbox
emr-4.8.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.8.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette de version Amazon EMR	Sqoop – Version d'environnement de test (sandbox)	Composants installés avec Sqoop-Sandbox
emr-4.7.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.7.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette de version Amazon EMR	Sqoop – Version d'environnement de test (sandbox)	Composants installés avec Sqoop-Sandbox
emr-4.7.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.6.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.6.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette de version Amazon EMR	Sqoop – Version d'environnement de test (sandbox)	Composants installés avec Sqoop-Sandbox
emr-4.5.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, sqoop-client
emr-4.4.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, sqoop-client

Zeppelin (versions de l'environnement de test)

Zeppelin est disponible sous la forme d'une application d'environnement de test (sandbox) à partir de la version Amazon EMR 4.1.0.

Informations sur la version Zeppelin-Sandbox

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.9.6	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
		server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.5	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.4	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.9.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.9.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.9.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.5	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.8.4	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.8.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.8.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.8.0	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.4	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.7.3	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.2	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.7.1	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.7.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.6.1	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.6.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.5.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.4.0	0.5.6	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette de version Amazon EMR	Version Zeppelin-Sandbox	Composants installés avec Zeppelin-Sandbox
emr-4.3.0	0.5.5	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.2.0	0.5.5	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-4.1.0	0.6.0-SNAPSHOT	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Zookeeper (versions de l'environnement de test)

Zookeeper est disponible sous la forme d'une application d'environnement de test (sandbox) à partir de la version Amazon EMR 4.6.0.

Informations sur la version ZooKeeper-Sandbox

Étiquette de version Amazon EMR	Zookeeper – Version d'environnement de test (sandbox)	Composants installés avec ZooKeeper-Sandbox
emr-4.9.6	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.9.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager

Étiquette de version Amazon EMR	Zookeeper – Version d'environnement de test (sandbox)	Composants installés avec ZooKeeper-Sandbox
		cemanager, zookeeper-client, zookeeper-server
emr-4.9.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, zookeeper-client, zookeeper-server
emr-4.9.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, zookeeper-client, zookeeper-server
emr-4.9.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcecemanager, zookeeper-client, zookeeper-server

Étiquette de version Amazon EMR	Zookeeper – Version d'environnement de test (sandbox)	Composants installés avec ZooKeeper-Sandbox
emr-4.8.5	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.8.4	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.8.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Étiquette de version Amazon EMR	Zookeeper – Version d'environnement de test (sandbox)	Composants installés avec ZooKeeper-Sandbox
emr-4.8.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.8.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Étiquette de version Amazon EMR	Zookeeper – Version d'environnement de test (sandbox)	Composants installés avec ZooKeeper-Sandbox
emr-4.7.4	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.7.3	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-4.7.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Étiquette de version Amazon EMR	Zookeeper – Version d'environnement de test (sandbox)	Composants installés avec ZooKeeper-Sandbox
emr-4.7.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.7.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-4.6.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Étiquette de version Amazon EMR	Zookeeper – Version d'environnement de test (sandbox)	Composants installés avec ZooKeeper-Sandbox
emr-4.6.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Considérations relatives à l'utilisation de Hive sur Amazon EMR 4.x

Cette section couvre les différences à prendre en compte lors de l'utilisation de la version Hive 1.0.0 sur les versions 4.x d'Amazon EMR, en comparaison avec Hive 2.x sur les versions 5.x d'Amazon EMR.

Les transactions ACID ne sont pas prises en charge.

Hive sur les versions 4.x d'Amazon EMR ne prend pas en charge les transactions ACID avec les données Hive stockées dans Amazon S3 lorsque vous utilisez les versions 4.x. Si vous essayez de créer une table transactionnelle dans Amazon S3, cela entraîne une exception.

Lecture et écriture de tables dans Amazon S3

Hive sur les versions 4.x d'Amazon EMR peut écrire directement dans Amazon S3 sans utiliser de fichiers temporaires. Cela améliore les performances, mais par voie de conséquence, vous ne pouvez pas lire et écrire dans la même table dans Amazon S3 via la même instruction Hive. Une solution de contournement consiste à créer et utiliser une table temporaire dans HDFS.

L'exemple suivant montre comment utiliser plusieurs instructions Hive pour mettre à jour une table dans Amazon S3. Les instructions créent une table temporaire nommée tmp dans HDFS en se basant sur une table dans Amazon S3 nommée my_s3_table. La table dans Amazon S3 est alors mise à jour avec le contenu de la table temporaire.

```
CREATE TEMPORARY TABLE tmp LIKE my_s3_table;
INSERT OVERWRITE TABLE tmp SELECT ....;
```

```
INSERT OVERWRITE TABLE my_s3_table SELECT * FROM tmp;
```

Log4j par rapport à Log4j 2

Hive sur les versions Amazon EMR 4.x utilise Log4j. À partir de la version 5.0.0, Log4j 2 est la configuration par défaut. Ces versions peuvent nécessiter différentes configurations de journalisation. Pour plus de détails, consultez [Apache Log4j 2](#).

MapReduce est le moteur d'exécution par défaut

Hive sur les versions Amazon EMR 4.x utilise MapReduce comme moteur d'exécution par défaut. À partir de la version Amazon EMR 5.0.0, Tez est la valeur par défaut, ce qui offre des performances améliorées pour la plupart des flux de travail.

Autorisation Hive

Hive sur les versions Amazon EMR 4.x prend en charge l'[autorisation Hive](#) pour HDFS, mais pas pour EMRFS et Amazon S3. Les clusters Amazon EMR s'exécutent avec l'autorisation désactivée par défaut.

Comportement de fusion de fichiers Hive avec Amazon S3

Hive sur les versions 4.x d'Amazon EMR fusionne de petits fichiers à la fin d'une tâche de mappage uniquement si `hive.merge.mapfiles` est défini sur `true`. Une fusion est déclenchée uniquement si la taille de sortie moyenne de la tâche est inférieure au paramètre `hive.merge.smallfiles.avgsize`. Le comportement d'Amazon EMR Hive est identique si le chemin de la sortie finale est sur HDFS. Cependant, si le chemin de sortie se trouve dans Amazon S3, le paramètre `hive.merge.smallfiles.avgsize` est ignoré. Dans ce cas, la tâche de fusion est toujours déclenchée si `hive.merge.mapfiles` est défini sur `true`.

Considérations relatives à l'utilisation de Pig sur Amazon EMR 4.x

La version Pig 0.14.0 est installée sur les clusters créés avec les versions Amazon EMR 4.x. Pig a été mis à niveau à la version 0.16.0 dans Amazon EMR 5.0.0. Les différences importantes sont discutées ci-dessous.

Moteur d'exécution par défaut différent

La version Pig 0.14.0 sur Amazon EMR 4.x utilise MapReduce comme moteur d'exécution par défaut. Pig 0.16.0 et les versions ultérieures utilisent Apache Tez. Vous pouvez explicitement définir

exectype=mapreduce dans la `pig-properties` classification de configuration pour utiliser MapReduce.

Fonctions Pig définies par l'utilisateur (UDF) abandonnées

Les fonctions définies par l'utilisateur qui étaient disponibles dans Pig sur les versions 4.x d'Amazon EMR ont été abandonnées à partir de la version Pig 0.16.0. La plupart des fonctions définies par l'utilisateur ont des fonctions équivalentes que vous pouvez utiliser à la place. Le tableau suivant répertorie les fonctions définies par l'utilisateur abandonnées et leur fonction équivalente. Pour plus d'informations, consultez [Fonctions intégrées](#) sur le site Apache Pig.

Fonctions définies par l'utilisateur abandonnées	Fonction équivalente
FORMAT_DT (formatdt, date)	GetHour(date), GetMinute(date), GetMonth(date), GetSecond(date), GetWeek(date), GetYear(date), GetDay(date)
EXTRACT(chaine, modèle)	REGEX_EXTRACT_ALL(chaine, modèle)
REPLACE(chaine, modèle, remplacement)	REPLACE(chaine, modèle, remplacement)
DATE_TIME()	ToDate()
DURATION(dt, dt2)	WeeksBetween(dt, dt2), YearsBetween(dt, dt2), SecondsBetween(dt, dt2), MonthsBetween(dt, dt2), MinutesBetween(dt, dt2), HoursBetween(dt, dt2)
EXTRACT_DT(format, date)	GetHour(date), GetMinute(date), GetMonth(date), GetSecond(date), GetWeek(date), GetYear(date), GetDay(date)
OFFSET_DT(date, durée)	AddDuration(date, durée), SubtractDuration(date, durée)
PERIOD(dt, dt2)	WeeksBetween(dt, dt2), YearsBetween(dt, dt2), SecondsBetween(dt, dt2), MonthsBetween(dt, dt2), MinutesBetween(dt, dt2), HoursBetween(dt, dt2)

Fonctions définies par l'utilisateur abandonnées	Fonction équivalente
CAPITALIZE(chaine)	UCFIRST(chaine)
CONCAT_WITH()	CONCAT()
INDEX_OF()	INDEXOF()
LAST_INDEX_OF()	LAST_INDEXOF()
SPLIT_ON_REGEX()	STRSPLT()
UNCAPITALIZE()	LCFIRST()

Les fonctions UDF suivantes ont été abandonnées sans équivalent : `FORMAT()`, `LOCAL_DATE()`, `LOCAL_TIME()`, `CENTER()`, `LEFT_PAD()`, `REPEAT()`, `REPLACE_ONCE()`, `RIGHT_PAD()`, `STRIP()`, `STRIP_END()`, `STRIP_START()`, `SWAP_CASE()`.

Commandes Grunt abandonnées

Certaines commandes Grunt ont été suspendues à commencer par Pig 0.16.0. Le tableau suivant répertorie les commandes Grunt sous Pig 0.14.0 et les commandes équivalentes dans la version actuelle, le cas échéant.

Commandes Grunt Pig 0.14.0 et commandes équivalentes de la version de Pig actuelle

Commande Grunt Pig 0.14.0	Commande Grunt sous Pig 0.16.0 et version ultérieure
<code>cat <non-hdfs-path></code>	<code>fs -cat <non-hdfs-path>;</code>
<code>cd <non-hdfs-path>;</code>	Pas d'équivalent
<code>ls <non-hdfs-path>;</code>	<code>fs -ls <non-hdfs-path>;</code>
<code>déplacer <non-hdfs-path> <non-hdfs-path>;</code>	<code>fs -mv <non-hdfs-path> <non-hdfs-path>;</code>
<code>copier <non-hdfs-path> <non-hdfs-path>;</code>	<code>fs -cp <non-hdfs-path> <non-hdfs-path>;</code>
<code>copyToLocal <non-hdfs-path> <local-path>;</code>	<code>fs -copyToLocal <non-hdfs-path> <local-path>;</code>

Commande Grunt Pig 0.14.0	Commande Grunt sous Pig 0.16.0 et version ultérieure
<code>copyFromLocal <local-path> <non-hdfs-path>;</code>	<code>fs -copyFromLocal <local-path> <non-hdfs-path>;</code>
<code>mkdir <non-hdfs-path>;</code>	<code>fs -mkdir <non-hdfs-path>;</code>
<code>rm <non-hdfs-path>;</code>	<code>fs -rm -r -skipTrash <non-hdfs-path>;</code>
<code>rmf <non-hdfs-path>;</code>	<code>fs -rm -r -skipTrash <non-hdfs-path>;</code>

Capacités supprimées pour les répertoires de base non HDFS

Pig 0.14.0 sur les versions 4.x d'Amazon EMR possède deux mécanismes pour autoriser les utilisateurs autres que l'utilisateur hadoop, qui ne disposent pas de répertoires de base, à exécuter des scripts Pig. Le premier mécanisme est un mécanisme de secours automatique qui définit le répertoire de travail initial dans le répertoire racine, si le répertoire de base n'existe pas. Le second mécanisme est une propriété `pig.initial.fs.name` qui vous permet de modifier le répertoire de travail initial.

Ces mécanismes ne sont pas disponibles à partir des versions Amazon EMR 5.0.0. C'est pourquoi les utilisateurs doivent disposer d'un répertoire de base sur un HDFS. Cela ne s'applique pas à l'utilisateur hadoop, car un répertoire de base est provisionné lors du lancement. Les scripts qui s'exécutent à l'aide des étapes JAR Hadoop utilisent par défaut l'utilisateur Hadoop, sauf si un autre utilisateur est explicitement spécifié à l'aide de `command-runner.jar`.

Amazon EMR version 4.9.6

versions de l'application 4.9.6

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

notes de mise à jour de la version 4.9.6

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

versions des composants 4.9.6

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.17.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-9	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	1.0.0-amzn-9	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-9	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-9	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.3	Clients de ligne de commande Spark.
spark-history-server	1.6.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.3	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.

Composant	Version	Description
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

4.9.6 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.9.6

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.

Classifications	Description
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.pr operties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole .properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra .properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.prop erties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.prope rties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.pro perties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile .properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.p roperties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.pro perties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresq l.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.pr operties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.pro perties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.prop erties de Presto.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.9.5

versions de l'application 4.9.5

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

notes de publication de la version 4.9.5

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 4.9.5 d'Amazon. Les modifications ont été apportées à la version 4.9.4.

Date de parution initiale : 29 août 2018

Modifications, améliorations et problèmes résolus

- HBase
 - Cette version corrige une vulnérabilité de sécurité potentielle.

versions des composants 4.9.5

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.17.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème

Composant	Version	Description
		ème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	1.0.0-amzn-9	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-9	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-9	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-9	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.

Composant	Version	Description
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.3	Clients de ligne de commande Spark.
spark-history-server	1.6.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	1.6.3	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

4.9.5 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.9.5

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.

Classifications	Description
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.9.4

versions de l'application 4.9.4

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

notes de publication de la version 4.9.4

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 4.9.4 d'Amazon. Les modifications ont été apportées à la version 4.9.3.

Date de parution initiale : 29 mars 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités potentielles.

versions des composants 4.9.4

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.17.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-9	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-9	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-9	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server	1.0.0-amzn-9	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.3	Clients de ligne de commande Spark.
spark-history-server	1.6.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.3	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

4.9.4 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.9.4

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase <code>hbase-log4j.properties</code> .

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog <code>proto-hive-site</code> fichier <code>.xml</code> .
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat <code>log4j.properties</code> de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier <code>webhcat-site.xml</code> de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier <code>hive-exec-log4j.properties</code> de Hive.

Classifications	Description
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.

Classifications	Description
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.

Classifications	Description
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.9.3

versions de l'application 4.9.3

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

notes de publication de la version 4.9.3

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.9.3 d'Amazon. Les modifications concernent la version EMR 4.9.2 d'Amazon.

Date de parution initiale : 22 janvier 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités associées à l'exécution spéculative (CVE-2017-5715, -2017-5753 et CVE-2017-5754). CVE Pour plus d'informations, consultez <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

versions des composants 4.9.3

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-`

composant avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
emr-ddb	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.17.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-9	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	1.0.0-amzn-9	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-9	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-9	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.3	Clients de ligne de commande Spark.
spark-history-server	1.6.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.3	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.

Composant	Version	Description
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

4.9.3 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.9.3

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.

Classifications	Description
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.pr operties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole .properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra .properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.prop erties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.prope rties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.pro perties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile .properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.p roperties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.pro perties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresq l.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.pr operties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.pro perties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.prop erties de Presto.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.9.2

versions de l'application 4.9.2

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

notes de publication de la version 4.9.2

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.9.2 d'Amazon. Les modifications concernent la version EMR 4.9.1 d'Amazon.

Date de parution : 13 juillet 2017

Des modifications mineures, des correctifs de bogues et des améliorations ont été apportées à cette version.

versions des composants 4.9.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon

EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.3.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.17.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
<code>ganglia-metadata-collector</code>	3.7.2	Collecteur de métadonnées Ganglia pour agréger les

Composant	Version	Description
		métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.

Composant	Version	Description
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-9	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.

Composant	Version	Description
hcatalog-server	1.0.0-amzn-9	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-9	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-9	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-9	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.

Composant	Version	Description
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.3	Clients de ligne de commande Spark.
spark-history-server	1.6.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.3	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	1.6.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

4.9.2 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.9.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.

Classifications	Description
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.9.1

versions de l'application 4.9.1

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.157.1
Spark	1.6.3	1.6.3	1.6.3	1.6.3
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.9

notes de publication de la version 4.9.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.9.1 d'Amazon. Les modifications concernent la version EMR 4.8.4 d'Amazon.

Date de parution : 10 avril 2017

Problèmes connus résolus depuis les versions précédentes

- [Rétroportages de HIVE-9976 et -10106 HIVE](#)
- Correction d'un problème à cause YARN duquel un grand nombre de nœuds (supérieur à 2 000) et de conteneurs (supérieur à 5 000) provoquait une out-of-memory erreur, par exemple : "Exception in thread main java.lang.OutOfMemoryError".

Modifications et améliorations

- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2017.03. Pour plus d'informations, consultez <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- Python 2.6 a été supprimé de l'image Linux EMR de base d'Amazon. Vous pouvez installer Python 2.6 manuellement si nécessaire.

versions des composants 4.9.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.3.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.15.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.3-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.3-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.7.3-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-9	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	1.0.0-amzn-9	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-9	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-9	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.3	Clients de ligne de commande Spark.
spark-history-server	1.6.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.3	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.

Composant	Version	Description
webserver	2.4.25+	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

4.9.1 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.9.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.

Classifications	Description
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.pr operties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole .properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra .properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.prop erties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.prope rties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.pro perties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile .properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.p roperties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.pro perties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresq l.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.pr operties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.pro perties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.prop erties de Presto.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.8.5

versions de l'application 4.8.5

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.3
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Presto-Sandbox	0.157.1	0.157.1	0.157.1	0.152.3
Spark	1.6.3	1.6.3	1.6.3	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.9	3.4.8

notes de publication de la version 4.8.5

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

Versions des composants 4.8.5

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.14.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème

Composant	Version	Description
		ème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	1.0.0-amzn-8	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-8	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-8	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-8	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-8	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.

Composant	Version	Description
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.3	Clients de ligne de commande Spark.
spark-history-server	1.6.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	1.6.3	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

4.8.5 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.8.5

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.

Classifications	Description
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.8.4

versions de l'application 4.8.4

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.3	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.157.1	0.152.3	0.151
Spark	1.6.3	1.6.3	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.9	3.4.8	3.4.8

notes de publication de la version 4.8.4

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.8.4 d'Amazon. Les modifications concernent la version EMR 4.8.3 d'Amazon.

Date de version : 7 février 2017

Des modifications mineures, des correctifs de bogues et des améliorations ont été apportées à cette version.

Versions des composants 4.8.4

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.14.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.7.3-amzn-1	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	1.0.0-amzn-8	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-8	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	1.0.0-amzn-8	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-8	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-8	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop.

Composant	Version	Description
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.54+	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.3	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	1.6.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.3	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	1.6.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.25+	HTTP serveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

4.8.4 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.8.4

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .

Classifications	Description
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive

Classifications	Description
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.

Classifications	Description
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-hbase.properties</code> de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier <code>hbase-site.xml</code> de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier <code>blackhole.properties</code> de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier <code>cassandra.properties</code> de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.

Classifications	Description
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.

Classifications	Description
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.8.3

versions de l'application 4.8.3

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.2
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.3	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.157.1	0.152.3	0.151	0.151
Spark	1.6.3	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.4
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0.6.1

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.8	3.4.8	3.4.8

notes de publication de la version 4.8.3

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.8.3 d'Amazon. Les modifications concernent la version EMR 4.8.2 d'Amazon.

Date de parution : 29 décembre 2016

Mises à niveau

- Mise à niveau vers Presto 0.157.1. Pour plus d'informations, consultez [Notes de mise à jour Presto](#) dans la documentation Presto.
- Mise à niveau vers Spark 1.6.3. Pour plus d'informations, consultez les [Notes de mise à jour Spark](#) dans la documentation Apache Spark.
- Mise à niveau vers la version ZooKeeper 3.4.9. Pour plus d'informations, consultez les [notes ZooKeeper de version](#) dans la ZooKeeper documentation d'Apache.

Modifications et améliorations

- Ajout de la prise en charge du type d'EC2instance Amazon m4.16xlarge dans les EMR versions 4.8.3 et ultérieures d'Amazon, à l'exception des versions 5.0.0, 5.0.3 et 5.2.0.
- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.09. Pour plus d'informations, consultez <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème dans Hadoop où le ReplicationMonitor thread pouvait rester bloqué pendant longtemps en raison d'une course entre la réplication et la suppression du même fichier dans un grand cluster.
- Correction d'un problème en raison duquel ControlledJob # toString échouait avec une exception de pointeur nul (NPE) lorsque le statut de la tâche n'était pas correctement mis à jour.

Versions des composants 4.8.3

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	4.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.13.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème

Composant	Version	Description
		ème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.3-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	1.0.0-amzn-8	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-8	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-8	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-8	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-8	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.

Composant	Version	Description
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.157.1	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.157.1	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.3	Clients de ligne de commande Spark.
spark-history-server	1.6.3	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	1.6.3	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.3	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.9	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.9	ZooKeeper client en ligne de commande.

4.8.3 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.8.3

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.

Classifications	Description
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.8.2

versions de l'application 4.8.2

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.2	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.3	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.152.3	0.151	0.151	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.4	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0.6.1	0,5,6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

notes de publication de la version 4.8.2

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.8.2 d'Amazon. Les modifications concernent la version EMR 4.8.0 d'Amazon.

Date de parution : 24 octobre 2016

Mises à niveau

- Mise à niveau vers Hadoop 2.7.3
- Mise à niveau vers Presto 0.152.3, qui comprend la prise en charge de l'interface Web Presto. Utilisez le port 8889 du coordinateur Presto pour accéder à l'interface Web Presto. Pour plus d'informations sur l'interface Web Presto, consultez [Interface Web](#) dans la documentation Presto.
- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.09. Pour plus d'informations, consultez <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Versions des composants 4.8.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	4.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-goodies	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.10.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.3-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.

Composant	Version	Description
hadoop-hdfs-library	2.7.3-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-ftpfs-server	2.7.3-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.3-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-7	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-7	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-7	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-7	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-7	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server	1.0.0-amzn-7	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.52	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.152.3	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.152.3	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.2	Clients de ligne de commande Spark.
spark-history-server	1.6.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

4.8.2 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.8.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase <code>hbase-log4j.properties</code> .

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.

Classifications	Description
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.

Classifications	Description
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.

Classifications	Description
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier <code>redis.properties</code> de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier <code>tpch.properties</code> de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier <code>spark-defaults.conf</code> de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.8.1

versions de l'application 4.8.1

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.2	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.151	0.151	0.148	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.4	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0.6.1	0,5.6	0,5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

notes de publication de la version 4.8.1

Versions des composants 4.8.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	3.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.9.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.2-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-4	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.7.2-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-7	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-7	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-7	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-7	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-7	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server	1.0.0-amzn-7	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.51	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.151	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.151	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.2	Clients de ligne de commande Spark.
spark-history-server	1.6.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

4.8.1 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

classifications EMR-4.8.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase <code>hbase-log4j.properties</code> .

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.

Classifications	Description
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.

Classifications	Description
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.

Classifications	Description
presto-connector-jmx	Modifiez les valeurs dans le fichier <code>jmx.properties</code> de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier <code>kafka.properties</code> de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier <code>localfile.properties</code> de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier <code>mongodb.properties</code> de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier <code>mysql.properties</code> de Presto.
presto-connector-postgresql	Modifiez les valeurs dans le fichier <code>postgresql.properties</code> de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier <code>raptor.properties</code> de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier <code>redis.properties</code> de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier <code>tpch.properties</code> de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier <code>spark-defaults.conf</code> de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Spark.

Classifications	Description
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraooop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.8.0

Versions de l'application 4.8.0

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.2	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.151	0.148	0.148	0.148
Spark	1.6.2	1.6.2	1.6.2	1.6.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.6.1	0,5.6	0,5.6	0,5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

notes de publication de la version 4.8.0

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.8.0 d'Amazon. Les modifications concernent la version EMR 4.7.2 d'Amazon.

Date de parution : 7 septembre 2016

Mises à niveau

- Mise à niveau vers la version HBase 1.2.2
- Mise à niveau vers Presto-Sandbox 0.151
- Mise à niveau vers Tez 0.8.4
- Mise à niveau vers Zeppelin-Sandbox 0.6.1

Modifications et améliorations

- Correction d'un problème en raison YARN duquel ils ApplicationMaster tentaient de nettoyer des conteneurs qui n'existent plus parce que leurs instances ont été résiliées.
- Corrigé les actions hive-server2 URL pour Hive2 dans les exemples d'Oozie.
- Ajout de la prise en charge d'autres catalogues Presto.
- [Correctifs rétroportés : HIVE-8948, HIVE -12679, -13405, -3116, HIVE -12689 PHOENIX HADOOP](#)
- Ajout de la prise en charge des configurations de sécurité, qui vous permettent de créer et d'appliquer des options de chiffrement plus facilement. Pour en savoir plus, consultez [Chiffrement des données](#).

Versions des composants 4.8.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	3.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.9.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
<code>ganglia-monitor</code>	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème

Composant	Version	Description
		ème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.2-amzn-4	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-4	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.2-amzn-4	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-4	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.2-amzn-4	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-4	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.2	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.2	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.2	HBaseclient en ligne de commande.
hbase-rest-server	1.2.2	Service fournissant un RESTful HTTP point de terminaison pourHBase.

Composant	Version	Description
hbase-thrift-server	1.2.2	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	1.0.0-amzn-7	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-7	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-7	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-7	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-7	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-7	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.

Composant	Version	Description
mysql-server	5.5.51	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.151	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.151	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.2	Clients de ligne de commande Spark.
spark-history-server	1.6.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.

Composant	Version	Description
spark-on-yarn	1.6.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.4	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.6.1	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

4.8.0 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.8.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.

Classifications	Description
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hiveserver2-site	Modifiez les valeurs dans le fichier hiveserver2-site.xml de Hive Server2
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-blackhole	Modifiez les valeurs dans le fichier blackhole.properties de Presto.
presto-connector-cassandra	Modifiez les valeurs dans le fichier cassandra.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
presto-connector-jmx	Modifiez les valeurs dans le fichier jmx.properties de Presto.
presto-connector-kafka	Modifiez les valeurs dans le fichier kafka.properties de Presto.
presto-connector-localfile	Modifiez les valeurs dans le fichier localfile.properties de Presto.
presto-connector-mongodb	Modifiez les valeurs dans le fichier mongodb.properties de Presto.
presto-connector-mysql	Modifiez les valeurs dans le fichier mysql.properties de Presto.

Classifications	Description
presto-connector-postgresql	Modifiez les valeurs dans le fichier postgresql.properties de Presto.
presto-connector-raptor	Modifiez les valeurs dans le fichier raptor.properties de Presto.
presto-connector-redis	Modifiez les valeurs dans le fichier redis.properties de Presto.
presto-connector-tpch	Modifiez les valeurs dans le fichier tpch.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.

Classifications	Description
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.7.4

versions de l'application 4.7.4

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.1
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
Mahout	0.12.2	0.12.2	0.12.2	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.148	0.148	0.147
Spark	1.6.2	1.6.2	1.6.2	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5.6	0,5.6	0,5.6	0,5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

notes de publication de la version 4.7.4

Il s'agit d'une version de correctif visant à ajouter l'authentification AWS Signature version 4 pour les demandes adressées à Amazon S3. Toutes les applications et tous les composants sont identiques à ceux de la EMR version précédente d'Amazon.

Important

Dans cette version, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. Pour de plus amples informations, veuillez consulter [Nouveautés](#).

versions des composants 4.7.4

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et sont installés pour les processus et les fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	3.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.8.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.2-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-3	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.7.2-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.

Composant	Version	Description
hbase-region-server	1.2.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.1	HBaseclient en ligne de commande.
hbase-rest-server	1.2.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-6	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-6	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-6	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-6	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-6	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server	1.0.0-amzn-6	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.46	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.148	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.148	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.2	Clients de ligne de commande Spark.
spark-history-server	1.6.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.3	L'YARN application Tez et les bibliothèques.
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.5.6-incubating	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.

Composant	Version	Description
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

4.7.4 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.7.4

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase <code>hbase-log4j.properties</code> .

Classifications	Description
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog <code>jndi.properties</code> .
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog <code>proto-hive-site</code> fichier <code>.xml</code> .
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat <code>log4j.properties</code> de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier <code>webhcat-site.xml</code> de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier <code>hive-exec-log4j.properties</code> de Hive.

Classifications	Description
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.

Classifications	Description
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.

Classifications	Description
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.7.3

versions de l'application 4.7.3

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,75
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.1

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.0	0.12.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	4.7.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.148	0.147	0.147
Spark	1.6.2	1.6.2	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5.6	0,5.6	0,5.6	0,5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

notes de publication de la version 4.7.3

versions des composants 4.7.3

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	3.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.8.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.2-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.2-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.1	HBaseclient en ligne de commande.
hbase-rest-server	1.2.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-6	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-6	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-6	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	1.0.0-amzn-6	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-6	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-6	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.46	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.148	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.148	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.2	Clients de ligne de commande Spark.
spark-history-server	1.6.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.3	L'YARN application Tez et les bibliothèques.

Composant	Version	Description
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.5.6-incubating	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

4.7.3 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

classifications EMR-4.7.3

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Classifications	Description
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.7.2

versions de l'application 4.7.2

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,75	1,1,27
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.1	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.2	0.12.0	0.12.0	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	4.7.0	-

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.148	0.147	0.147	0.143
Spark	1.6.2	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	0.8.3	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5.6	0,5.6	0,5.6	0,5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

notes de publication de la version 4.7.2

Les notes de mise à jour suivantes contiennent des informations relatives à Amazon EMR 4.7.2.

Date de parution : 15 juillet 2016

Fonctionnalités

- Mise à niveau vers Mahout 0.12.2
- Mise à niveau vers Presto 0.148

- Mise à niveau vers Spark 1.6.2
- Vous pouvez désormais créer un `AWSCredentialsProvider` à utiliser en EMRFS utilisant a URI comme paramètre. Pour plus d'informations, consultez la section [Créer un `AWSCredentialsProvider` formulaire EMRFS](#).
- EMRFS permet désormais aux utilisateurs de configurer un point de terminaison DynamoDB personnalisé pour leurs métadonnées Consistent View à l'aide `fs.s3.consistent.dynamodb.endpoint` de la propriété `in.emrfs-site.xml`
- Ajout d'un script dans `/usr/bin` appelé `spark-example`, qui encapsule `/usr/lib/spark/spark/bin/run-example` pour vous permettre d'exécuter des exemples directement. Par exemple, pour exécuter l' SparkPi exemple fourni avec la distribution Spark, vous pouvez l'exécuter `spark-example SparkPi 100` depuis la ligne de commande ou en utilisant `command-runner.jar` comme étape leAPI.

Problèmes connus résolus depuis les versions précédentes

- Résolution d'un problème où `spark-assembly.jar` pour Oozie n'était pas à l'emplacement approprié quand Spark était également installé, ce qui provoquait un échec du lancement d'applications Spark avec Oozie.
- Correction d'un problème lié à la journalisation dans les conteneurs basée sur Spark Log4J. YARN

versions des composants 4.7.2

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	3.2.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.1.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.8.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.2-amzn-3	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-3	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.2-amzn-3	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-3	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.7.2-amzn-3	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.1	HBaseclient en ligne de commande.
hbase-rest-server	1.2.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-6	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-6	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-6	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	1.0.0-amzn-6	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-6	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-6	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.2	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.46	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.148	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.148	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.2	Clients de ligne de commande Spark.
spark-history-server	1.6.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.3	L'YARN application Tez et les bibliothèques.

Composant	Version	Description
webserver	2.4.23	HTTPserveur Apache.
zeppelin-server	0.5.6-incubating	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

4.7.2 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.7.2

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hadoop-ssl-server	Modifier la configuration du serveur ssl hadoop
hadoop-ssl-client	Modifier la configuration du client ssl hadoop
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.

Classifications	Description
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.

Classifications	Description
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Classifications	Description
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.7.1

versions de l'application 4.7.1

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
Kit AWS SDK pour Java	1,1,75	1,1,75	1,1,27	1,1,27
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.1	1.2.0	1.2.0
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.0	0.12.0	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	4.7.0	-	-

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.147	0.147	0.143	0.143
Spark	1.6.1	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	0.8.3	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,6	0,5,6	0,5,6	0,5,6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	3.4.8

notes de publication de la version 4.7.1

Les notes de mise à jour suivantes contiennent des informations relatives à Amazon EMR 4.7.1.

Date de parution : 10 juin 2016

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème qui prolongeait le temps de démarrage des clusters lancés dans un environnement VPC doté de sous-réseaux privés. Le bogue ne concernait que les clusters lancés avec la version EMR 4.7.0 d'Amazon.
- Correction d'un problème qui entraînait une gestion incorrecte de la liste des fichiers dans Amazon EMR pour les clusters lancés avec la EMR version 4.7.0 d'Amazon.

Versions des composants 4.7.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	3.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.7.1	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.2-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.

Composant	Version	Description
hadoop-https-server	2.7.2-amzn-2	HTTP point de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARN service de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-2	YARN service d'allocation et de gestion des ressources du cluster et des applications distribuées.
hadoop-yarn-timeline-server	2.7.2-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.1	HBase client en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.2.1	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.2.1	Service fournissant un point de terminaison Thrift pour HBase
hcatalog-client	1.0.0-amzn-5	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-5	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	1.0.0-amzn-5	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-5	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-5	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop

Composant	Version	Description
mahout-client	0.12.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.46	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.147	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.147	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.1	Clients de ligne de commande Spark.

Composant	Version	Description
spark-history-server	1.6.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	1.6.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.3	L'YARN application Tez et les bibliothèques.
webserver	2.4.18	HTTP serveur Apache.
zeppelin-server	0.5.6-incubating	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

4.7.1 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.7.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.

Classifications	Description
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.

Classifications	Description
phoenix-metrics	Modifiez les valeurs dans le fichier <code>hadoop-metrics2-phoenix.properties</code> de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier <code>pig.properties</code> de Pig.
pig-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Pig.
presto-log	Modifiez les valeurs dans le fichier <code>log.properties</code> de Presto.
presto-config	Modifiez les valeurs dans le fichier <code>config.properties</code> de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier <code>hive.properties</code> de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier <code>spark-defaults.conf</code> de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Spark.
spark-metrics	Modifiez les valeurs dans le fichier <code>metrics.properties</code> de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier <code>oraoop-site.xml</code> de OraOop Sqoop.

Classifications	Description
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.7.0

Versions de l'application 4.7.0

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)

- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
Kit AWS SDK pour Java	1,1,75	1,1,27	1,1,27	1,1,27
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.1	1.2.0	1.2.0	-
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.2
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.12.0	0.11.1	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	4.7.0	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.147	0.143	0.143	0.140
Spark	1.6.1	1.6.1	1.6.1	1.6.1
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	0.8.3	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,6	0,5,6	0,5,6	0,5,6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	3.4.8	-

notes de publication de la version 4.7.0

Important

Amazon EMR 4.7.0 est obsolète. Utilisez plutôt Amazon EMR 4.7.1 ou version ultérieure.

Date de parution : 2 juin 2016

Fonctionnalités

- Ajout d'Apache Phoenix 4.7.0
- Ajout d'Apache Tez 0.8.3
- Mise à niveau vers la version HBase 1.2.1
- Mise à niveau vers Mahout 0.12.0
- Mise à niveau vers Presto 0.147
- Mise à niveau de la AWS SDK for Java version 1.10.75
- L'indicateur final a été supprimé de la propriété `mapreduce.cluster.local.dir` dans `mapred-site.xml` pour permettre aux utilisateurs d'exécuter Pig en mode local.
- Les JDBC pilotes Amazon Redshift sont disponibles sur le cluster

Les JDBC pilotes Amazon Redshift sont désormais inclus sur `/usr/share/aws/redshift/jdbc` `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` est le pilote Amazon JDBC Redshift compatible 4.1 `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` et le pilote Amazon Redshift JDBC compatible 4.0. Pour plus d'informations, consultez [Configurer une JDBC connexion](#) dans le guide de gestion Amazon Redshift.

- Java 8

À l'exception de Presto, Open JDK 1.7 est la version par défaut JDK utilisée pour toutes les applications. Cependant, Open JDK 1.7 et 1.8 sont tous deux installés. Pour plus d'informations sur la configuration de `JAVA_HOME` pour les applications, consultez [Configuration d'applications pour utiliser Java 8](#).

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème de noyau qui affectait de manière significative les performances des EBS volumes optimisés pour le débit HDD (st1) pour Amazon EMR dans `emr-4.6.0`.

- Correction d'un problème en raison duquel un cluster échouait si une zone de HDFS chiffrement était spécifiée sans choisir Hadoop comme application.
- La politique d'HDFSécriture par défaut est passée de RoundRobin àAvailableSpaceVolumeChoosingPolicy. Certains volumes n'ont pas été correctement utilisés avec la RoundRobin configuration, ce qui a entraîné la défaillance des nœuds principaux et un manque de fiabilitéHDFS.
- Correction d'un problème lié au EMRFSLI, qui provoquait une exception lors de la création de la table de métadonnées DynamoDB par défaut pour des vues cohérentes.
- Correction d'un problème de blocage EMRFS qui se produisait potentiellement lors d'opérations de renommage et de copie en plusieurs parties.
- Correction d'un problème en EMRFS raison duquel la CopyPart taille par défaut était de 5 Mo. La valeur par défaut est maintenant correctement définie sur 128 Mo.
- Résolution d'un problème lié à la configuration upstart dans Zeppelin qui pouvait vous empêcher potentiellement d'arrêter le service.
- Correction d'un problème lié à Spark et Zeppelin, qui vous empêchait d'utiliser le s3a:// URI schéma car il n'usr/lib/hadoop/hadoop-aws.jarétait pas correctement chargé dans leur chemin de classe respectif.
- Rétroporté [HUE-2484](#).
- J'ai rétroporté un [commit](#) depuis Hue 3.9.0 (JIRAil n'existe pas) pour corriger un problème avec l'exemple de HBase navigateur.
- Rétroporté [HIVE-9073](#).

Versions des composants 4.7.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par emr ou aws. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-

composant avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
emr-ddb	3.1.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.2.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.4.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.7.1	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.2-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.2-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hadoop-yarn-timeline-server	2.7.2-amzn-2	Service de récupération des informations actuelles et historiques pour les YARN applications.
hbase-hmaster	1.2.1	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.1	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.1	HBaseclient en ligne de commande.
hbase-rest-server	1.2.1	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.1	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-5	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-5	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-5	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	1.0.0-amzn-5	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-5	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-5	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-7	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.12.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5.46	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
phoenix-library	4,7,0- -1,2 HBase	Les bibliothèques phoenix pour serveur et client

Composant	Version	Description
phoenix-query-server	4,7,0- -1,2 HBase	Un serveur léger fournissant un JDBC accès ainsi que des tampons de protocole et un accès JSON au format à l'Avatica API
presto-coordinator	0.147	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.147	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.1	Clients de ligne de commande Spark.
spark-history-server	1.6.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
tez-on-yarn	0.8.3	L'YARN application Tez et les bibliothèques.

Composant	Version	Description
webserver	2.4.18	HTTPserveur Apache.
zeppelin-server	0.5.6-incubating	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

4.7.0 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.7.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.

Classifications	Description
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.

Classifications	Description
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log 4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
phoenix-hbase-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-hbase.properties de Phoenix.
phoenix-hbase-site	Modifiez les valeurs dans le fichier hbase-site.xml de Phoenix.
phoenix-log4j	Modifiez les valeurs dans le fichier log4j.properties de Phoenix.
phoenix-metrics	Modifiez les valeurs dans le fichier hadoop-metrics2-phoenix.properties de Phoenix.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
tez-site	Modifiez les valeurs dans le fichier tez-site.xml de Tez.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.

Classifications	Description
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.6.1

versions de l'application 4.6.1

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
Kit AWS SDK pour Java	1,1,27	1,1,27	1,1,27	1,1,27
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	1.2.0	-	-
HCatalog	1.0.0	1.0.0	1.0.0	1.0.0
Hadoop	2.7.2	2.7.2	2.7.2	2.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.1	0.11.1
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
Presto-Sandbox	0.143	0.143	0.140	0.136
Spark	1.6.1	1.6.1	1.6.1	1.6.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	1.4.6
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5.6	0,5.6	0,5.6	0,5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	-	-

notes de publication de la version 4.6.1

Versions des composants 4.6.1

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion`

commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
emr-ddb	3.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.1.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.3.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.6.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par

Composant	Version	Description
		le collecteur de métadonnées Ganglia.
hadoop-client	2.7.2-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.2-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.7.2-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hbase-hmaster	1.2.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.0	HBaseclient en ligne de commande.
hbase-rest-server	1.2.0	Service fournissant un RESTful HTTP point de terminaison pourHBase.
hbase-thrift-server	1.2.0	Service fournissant un point de terminaison Thrift pour. HBase
hcatalog-client	1.0.0-amzn-4	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-4	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.

Composant	Version	Description
hive-client	1.0.0-amzn-4	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-4	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-4	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-6	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.11.1	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
presto-coordinator	0.143	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.143	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.1	Clients de ligne de commande Spark.
spark-history-server	1.6.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
webserver	2,4	HTTPserveur Apache.
zeppelin-server	0.5.6-incubating	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

4.6.1 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

classifications EMR-4.6.1

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Modifiez les valeurs dans le fichier HBase <code>hadoop-metrics2-hbase.properties</code> .
hbase-policy	Modifiez les valeurs dans HBase le fichier <code>hbase-policy.xml</code> .
hbase-site	Modifiez les valeurs dans HBase le fichier <code>hbase-site.xml</code> .
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.

Classifications	Description
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.

Classifications	Description
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.

Classifications	Description
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Classifications	Description
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.6.0

Versions de l'application 4.6.0

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Zeppelin-Sandbox](#) et [ZooKeeper-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
Kit AWS SDK pour Java	1,1,27	1,1,27	1,1,27	1,1,27
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.7.2
HBase	1.2.0	-	-	-
HCatalog	1.0.0	1.0.0	1.0.0	-
Hadoop	2.7.2	2.7.2	2.7.1	2.7.1
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.1	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.143	0.140	0.136	0.130
Spark	1.6.1	1.6.1	1.6.0	1.6.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	1.4.6	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,6	0,5,6	0,5,6	0,5,5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	-	-	-

notes de publication de la version 4.6.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 4.6.0 d'Amazon.

- Ajouté HBase 1.2.0
- Ajout de Zookeeper-Sandbox 3.4.8
- Mise à niveau vers Presto-Sandbox 0.143

- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.03.0. Pour plus d'informations, consultez <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.
- Problème affectant les types de volumes optimisés pour le débit HDD (st1) EBS

Un problème lié aux versions 4.2 et supérieures du noyau Linux affecte de manière significative les performances des EBS volumes optimisés pour le débit HDD (st1) pour EMR. Cette version (emr-4.6.0) utilise la version de noyau 4.4.5 et est donc concernée. Par conséquent, nous vous recommandons de ne pas utiliser emr-4.6.0 si vous souhaitez utiliser des volumes st1. EBS Vous pouvez utiliser emr-4.5.0 ou des EMR versions antérieures d'Amazon avec st1 sans impact. En outre, nous fournirons le correctif avec les versions futures.

- Version de Python par défaut

Python 3.4 est maintenant installé par défaut, mais Python 2.7 reste la valeur système par défaut. Vous pouvez configurer Python 3.4 comme valeur par défaut du système à l'aide d'une action bootstrap ; vous pouvez utiliser la configuration API pour définir `PYSPARK_PYTHON` export to `/usr/bin/python3.4` dans la `spark-env` classification afin d'affecter la version de Python utilisée par PySpark.

- Java 8

À l'exception de Presto, Open JDK 1.7 est la version par défaut JDK utilisée pour toutes les applications. Cependant, Open JDK 1.7 et 1.8 sont tous deux installés. Pour plus d'informations sur la configuration de `JAVA_HOME` pour les applications, consultez [Configuration d'applications pour utiliser Java 8](#).

Problèmes connus résolus depuis les versions précédentes

- Résolution d'un problème où le provisionnement d'application échouait parfois de manière aléatoire en raison d'un mot de passe généré.
- Auparavant, `mysqld` était installé sur tous les nœuds. Maintenant, il est uniquement installé sur l'instance principale et seulement si l'application choisie inclut `mysql-server` comme composant. Actuellement, les applications suivantes incluent le `mysql-server` composant : HiveHCatalog, Hue, Presto-Sandbox et Sqoop-Sandbox.
- La valeur par défaut est passée `yarn.scheduler.maximum-allocation-vcores` à 80 par rapport à 32, ce qui corrige un problème introduit dans emr-4.4.0 qui se produit principalement avec Spark lors de l'utilisation de `maximizeResourceAllocation` option dans un cluster dont le type d'instance principal est l'un des rares types d'instance dont le nombre de YARN cœurs est

supérieur à 32, à savoir c4.8xlarge, d2.8xlarge, d2.8xlarge ou m4.10xlarge ont été concernés par ce problème.

- s3-dist-cp est désormais utilisé pour toutes les nominations EMRFS Amazon S3 et n'est plus transféré vers un répertoire temporaire. HDFS
- Résolution d'un problème lié au traitement des exceptions pour le chiffrement côté client du chargement partitionné.
- Ajout d'une option pour permettre aux utilisateurs de modifier la classe de stockage Amazon S3. Par défaut, ce paramètre est STANDARD. La configuration de classification `emrfs-site` est `fs.s3.storageClass`, et les valeurs possibles sont STANDARD, STANDARD_IA et REDUCED_REDUNDANCY. Pour plus d'informations sur les classes de stockage, consultez la section [Classes de stockage](#) dans le Guide de l'utilisateur Amazon Simple Storage Service.

Versions des composants 4.6.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	3.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.

Composant	Version	Description
emr-kinesis	3.1.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.3.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.6.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.2-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-1	HDFSclient et bibliothèque en ligne de commande

Composant	Version	Description
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.7.2-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.2-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hbase-hmaster	1.2.0	Service pour un HBase cluster chargé de la coordination des régions et de l'exécution des commandes administratives.
hbase-region-server	1.2.0	Service destiné à desservir une ou plusieurs HBase régions.
hbase-client	1.2.0	HBaseclient en ligne de commande.

Composant	Version	Description
hbase-rest-server	1.2.0	Service fournissant un RESTful HTTP point de terminaison pour HBase.
hbase-thrift-server	1.2.0	Service fournissant un point de terminaison Thrift pour HBase.
hcatalog-client	1.0.0-amzn-4	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-4	Fourniture de services HCatalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP point de terminaison fournissant une REST interface pour HCatalog.
hive-client	1.0.0-amzn-4	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-4	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-4	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-6	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop.

Composant	Version	Description
mahout-client	0.11.1	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
presto-coordinator	0.143	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.143	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.1	Clients de ligne de commande Spark.
spark-history-server	1.6.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.1	Moteur d'exécution en mémoire pour. YARN

Composant	Version	Description
spark-yarn-slave	1.6.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
webserver	2,4	HTTPserveur Apache.
zeppelin-server	0.5.6-incubating	Portable basé sur le web qui permet l'analyse de données interactives.
zookeeper-server	3.4.8	Service centralisé conçu pour la conservation des informations de configuration, l'affectation de noms, la synchronisation distribuée et la fourniture de services de groupe.
zookeeper-client	3.4.8	ZooKeeper client en ligne de commande.

Classifications de configuration 4.6.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.6.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.

Classifications	Description
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hbase-env	Changez les valeurs dans HBase l'environnement.
hbase-log4j	Modifiez les valeurs dans le fichier HBase hbase-log4j.properties.
hbase-metrics	Modifiez les valeurs dans le fichier HBase hadoop-metrics2-hbase.properties.
hbase-policy	Modifiez les valeurs dans HBase le fichier hbase-policy.xml.
hbase-site	Modifiez les valeurs dans HBase le fichier hbase-site.xml.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.

Classifications	Description
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS

Classifications	Description
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.

Classifications	Description
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.
zookeeper-config	Modifiez les valeurs dans ZooKeeper le fichier zoo.cfg.
zookeeper-log4j	Modifiez les valeurs dans le ZooKeeper fichier log4j.properties.

Amazon EMR version 4.5.0

Versions de l'application 4.5.0

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) et [Zeppelin-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
Kit AWS SDK pour Java	1,1,27	1,1,27	1,1,27	1,1,27
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.6.0
HBase	-	-	-	-

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
HCatalog	1.0.0	1.0.0	-	-
Hadoop	2.7.2	2.7.1	2.7.1	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.1	0.11.1	0.11.0	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.2.0
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.140	0.136	0.130	0.125
Spark	1.6.1	1.6.0	1.6.0	1.5.2
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	1.4.6	-	-

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5.6	0,5.6	0,5,5	0,5,5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

notes de publication de la version 4.5.0

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.5.0 d'Amazon.

Date de parution : 4 avril 2016

Fonctionnalités

- Mise à niveau vers Spark 1.6.1
- Mise à niveau vers Hadoop 2.7.2
- Mise à niveau vers Presto 0.140
- Ajout de la AWS KMS prise en charge du chiffrement côté serveur Amazon S3.

Problèmes connus résolus depuis les versions précédentes

- Problème résolu : les serveurs My SQL et Apache ne démarraient pas après le redémarrage d'un nœud.

- Correction d'un problème en raison duquel les tables non partitionnées stockées dans Amazon S3 IMPORT ne fonctionnaient pas correctement
- Résolution d'un problème lié à Presto où le répertoire intermédiaire devait être /mnt/tmp au lieu de /tmp lors de l'écriture dans des tables Hive.

Versions des composants 4.5.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	3.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.1.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.2.0	Application de copie distribuée optimisée pour Amazon S3.

Composant	Version	Description
emrfs	2.5.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.2-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.2-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.2-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.2-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.2-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.

Composant	Version	Description
hadoop-kms-server	2.7.2-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.2-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.2-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hcatalog-client	1.0.0-amzn-4	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-4	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTPpoint de terminaison fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-4	Client de ligne de commande Hive.

Composant	Version	Description
hive-metastore-server	1.0.0-amzn-4	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-4	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-5	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.11.1	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
presto-coordinator	0.140	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.140	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.

Composant	Version	Description
spark-client	1.6.1	Clients de ligne de commande Spark.
spark-history-server	1.6.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.1	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.6.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
webserver	2,4	HTTPserveur Apache.
zeppelin-server	0.5.6-incubating	Portable basé sur le web qui permet l'analyse de données interactives.

4.5.0 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.5.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.

Classifications	Description
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.

Classifications	Description
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.

Classifications	Description
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.

Classifications	Description
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Amazon EMR version 4.4.0

Versions de l'application 4.4.0

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) et [Zeppelin-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
Kit AWS SDK pour Java	1,1,27	1,1,27	1,1,27	Non suivie
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.6.0	-
HBase	-	-	-	-
HCatalog	1.0.0	-	-	-
Hadoop	2.7.1	2.7.1	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	3.7.1
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
Mahout	0.11.1	0.11.0	0.11.0	0.11.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.2.0	4.0.1
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.136	0.130	0.125	0.119
Spark	1.6.0	1.6.0	1.5.2	1.5.0
Sqoop	-	-	-	-
Sqoop-Sandbox	1.4.6	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,6	0,5,5	0,5,5	0,6,0-SNAPSHOT
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

notes de publication de la version 4.4.0

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.4.0 d'Amazon.

Date de parution : 14 mars 2016

Fonctionnalités

- Ajouté HCatalog 1.0.0
- Ajout de Sqoop-Sandbox 1.4.6
- Mise à niveau vers Presto 0.136
- Mise à niveau vers Zeppelin 0.5.6
- Mise à niveau vers Mahout 0.11.1
- `dynamicResourceAllocation` activé par défaut.
- Ajout d'un tableau de toutes les classifications de configuration pour la version. Pour en savoir plus, consultez le tableau Classifications des configurations dans [Configuration des applications](#).

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème à cause duquel le `maximizeResourceAllocation` paramètre ne réservait pas suffisamment de mémoire aux YARN ApplicationMaster démons.
- Correction d'un problème rencontré avec une personnalisationDNS. Si des entrées dans `resolve.conf` précèdent les entrées personnalisées fournies, les entrées personnalisées ne peuvent pas être résolues. Ce comportement a été affecté par les clusters dans VPC lesquels le serveur de VPC noms par défaut est inséré en tant qu'entrée supérieure `resolve.conf`.
- Résolution d'un problème où la version Python par défaut passait à la version 2.7 et boto n'était pas installé pour cette version.
- Correction d'un problème à cause duquel les YARN conteneurs et les applications Spark généraient un fichier de base de données Ganglia Round Robin (rrd) unique, ce qui entraînait le remplissage du premier disque attaché à l'instance. En raison de ce correctif, les métriques au niveau du YARN conteneur ont été désactivées et les métriques au niveau de l'application Spark ont été désactivées.
- Résolution d'un problème dans le transmetteur de journaux où celui-ci supprimait tous les dossiers de journal vides. Cela a eu pour effet que la Hive n'CLl a pas pu se connecter car Log Pusher supprimait le `user` dossier vide situé en dessous. `/var/log/hive`

- Résolution d'un problème affectant les importations Hive qui avait une incidence sur le partitionnement et entraînait une erreur lors de l'importation.
- Correction d'un problème à cause duquel EMRFS s3-dist-cp ne gérait pas correctement les noms de compartiments contenant des points.
- Modification d'un comportement de EMRFS telle sorte que dans les compartiments compatibles avec les versions, le fichier de `_$folder$` marqueurs n'est pas créé en permanence, ce qui peut contribuer à améliorer les performances des compartiments compatibles avec les versions.
- Le comportement a été modifié de EMRFS telle sorte qu'il n'utilise pas de fichiers d'instructions, sauf dans les cas où le chiffrement côté client est activé. Si vous souhaitez supprimer des fichiers d'instruction tout en utilisant le chiffrement côté client, vous pouvez définir la propriété `emrfs-site.xml fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled` sur `true`.
- Modification de l'agrégation des YARN journaux pour conserver les journaux à la destination de l'agrégation pendant deux jours. La destination par défaut est votre HDFS stockage en cluster. Si vous souhaitez modifier cette durée, remplacez la valeur `yarn.log-aggregation.retain-seconds` à l'aide de la classification de configuration `yarn-site` lorsque vous créez votre cluster. Comme toujours, vous pouvez enregistrer vos journaux d'applications dans Amazon S3 à l'aide du paramètre `log-uri` lorsque vous créez votre cluster.

Correctifs appliqués

- [HIVE-9655](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

Versions des composants 4.4.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion.EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-`

composant avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
emr-ddb	3.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.1.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.2.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.4.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.7.1-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.1-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.1-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.1-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.7.1-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.1-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.7.1-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.1-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.1-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hcatalog-client	1.0.0-amzn-3	Client de ligne de commande « hcat » pour la manipulation de hcatalog-server.
hcatalog-server	1.0.0-amzn-3	Fourniture de servicesH Catalog, table et couche de gestion du stockage pour les applications distribuées.
hcatalog-webhcat-server	1.0.0-amzn-3	HTTPpoint de terminais on fournissant une REST interface pourHCatalog.
hive-client	1.0.0-amzn-3	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-3	Service d'accès au métastore Hive, un référentiel sémantique e stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-3	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-5	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.11.1	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5	Mon serveur SQL de base de données.

Composant	Version	Description
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
presto-coordinator	0.136	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.136	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.0	Clients de ligne de commande Spark.
spark-history-server	1.6.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	1.6.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
sqoop-client	1.4.6	Client de ligne de commande Apache Sqoop.
webserver	2,4	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0.5.6-incubating	Portable basé sur le web qui permet l'analyse de données interactives.

4.4.0 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.4.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hcatalog-env	Changez les valeurs dans HCatalog l'environnement.

Classifications	Description
hcatalog-server-jndi	Modifiez les valeurs dans le fichier HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Modifiez les valeurs dans le HCatalog proto-hive-site fichier .xml.
hcatalog-webhcat-env	Modifiez les valeurs dans ebHCat l'environnement de HCatalog W.
hcatalog-webhcat-log4j	Modifiez les valeurs dans ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Modifiez les valeurs dans ebHCat le fichier webhcat-site.xml de HCatalog W.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.

Classifications	Description
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
sqoop-env	Modifiez les valeurs d'environnement Sqoop.
sqoop-oraoop-site	Modifiez les valeurs dans le fichier oraoop-site.xml de OraOop Sqoop.
sqoop-site	Modifiez les valeurs dans le fichier sqoop-site.xml de Sqoop.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Amazon EMR version 4.3.0

Versions de l'application 4.3.0

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) et [Zeppelin-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Kit AWS SDK pour Java	1,1,27	1,1,27	Non suivie	Non suivie
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,5	0,5,5	0,6,0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

notes de publication de la version 4.3.0

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.3.0 d'Amazon.

Date de parution : 19 janvier 2016

Fonctionnalités

- Mise à niveau vers Hadoop 2.7.1
- Mise à niveau vers Spark 1.6.0
- Mise à niveau de Ganglia vers la version 3.7.2
- Mise à niveau vers de Presto vers la version 0.130
- Amazon EMR a apporté quelques modifications au `spark.dynamicAllocation.enabled` moment où il est défini sur `true` ; il est défini sur `false` par défaut. Lorsque ce paramètre est défini sur `true`, cela a un effet sur les valeurs par défaut définies par le paramètre `maximizeResourceAllocation`.
 - Si `spark.dynamicAllocation.enabled` est défini sur `true`, `spark.executor.instances` n'est pas défini par `maximizeResourceAllocation`.

- Le paramètre `spark.driver.memory` est désormais configuré en fonction des types d'instances du cluster d'une manière similaire à la façon dont `spark.executors.memory` est défini. Cependant, étant donné que l'application du pilote Spark peut s'exécuter sur l'instance principale ou sur l'une des instances principales (par exemple, en mode YARN client et en mode cluster, respectivement), le `spark.driver.memory` paramètre est défini en fonction du type d'instance du type d'instance le plus petit entre ces deux groupes d'instances.
- Le `spark.default.parallelism` paramètre est désormais défini sur le double du nombre de CPU cœurs disponibles pour les YARN conteneurs. Dans les versions précédentes, c'était la moitié de cette valeur.
- Les calculs de la surcharge de mémoire réservée aux YARN processus Spark ont été ajustés pour être plus précis, ce qui a entraîné une légère augmentation de la quantité totale de mémoire disponible pour Spark (c'est-à-dire, `spark.executor.memory`).

Problèmes connus résolus depuis les versions précédentes

- YARN l'agrégation des journaux est désormais activée par défaut.
- Problème résolu : les journaux n'étaient pas transférés vers le compartiment de journaux Amazon S3 pour le cluster lorsque l'agrégation des YARN journaux était activée.
- YARN les tailles de conteneur ont désormais un nouveau minimum de 32 pour tous les types de nœuds.
- Résolution d'un problème lié à Ganglia qui entraînait des E/S de disque excessives sur le nœud primaire dans les clusters de grande taille.
- Résolution d'un problème qui empêchait les journaux des applications d'être transmis à Amazon S3 lorsqu'un cluster s'arrêtait.
- Correction d'un problème EMRFS CLI qui provoquait l'échec de certaines commandes.
- Correction d'un problème lié à Zeppelin qui empêchait le chargement des dépendances dans le sous-jacent. SparkContext
- Résolution d'un problème provoqué par une tentative de redimensionnement pour ajouter des instances.
- Correction d'un problème dans Hive à cause duquel `CREATE TABLE AS SELECT` envoyait trop d'appels de liste à Amazon S3.
- Résolution d'un problème où les grands clusters n'étaient pas provisionnés correctement lorsque Hue, Oozie et Ganglia étaient installés.

- Résolution d'un problème dans s3-dist-cp où un code de sortie zéro était renvoyé même en cas d'échec avec une erreur.

Versions des composants 4.3.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	3.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.1.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.1.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.3.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
ganglia-monitor	3.7.2	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.7.2	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.7.1	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.
hadoop-client	2.7.1-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.7.1-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.7.1-amzn-0	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.7.1-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.7.1-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.7.1-amzn-0	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API

Composant	Version	Description
hadoop-mapred	2.7.1-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.7.1-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.7.1-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hive-client	1.0.0-amzn-2	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-2	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-2	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-5	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.11.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5	Mon serveur SQL de base de données.

Composant	Version	Description
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
presto-coordinator	0.130	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.130	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.6.0	Clients de ligne de commande Spark.
spark-history-server	1.6.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.6.0	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	1.6.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
webserver	2,4	HTTPserveur Apache.

Composant	Version	Description
zeppelin-server	0.5.5-incubating-amzn-1	Portable basé sur le web qui permet l'analyse de données interactives.

4.3.0 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.3.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.
core-site	Modifiez les valeurs dans le fichier <code>core-site.xml</code> de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier <code>log4j.properties</code> de Hadoop.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier <code>hdfs-site.xml</code> .
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier <code>hive-exec-log4j.properties</code> de Hive.

Classifications	Description
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFSenvironment.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.

Classifications	Description
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Classifications	Description
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Amazon EMR version 4.2.0

Versions de l'application 4.2.0

Les applications suivantes sont prises en charge dans cette version : [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) et [Zeppelin-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Kit AWS SDK pour Java	1,1,27	1,1,27	Non suivie	Non suivie
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,5	0,5,5	0,6,0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

notes de publication de la version 4.2.0

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.2.0 d'Amazon.

Date de parution : 18 novembre 2015

Fonctionnalités

- Ajout de la prise en charge de Ganglia
- Mise à niveau vers Spark 1.5.2
- Mise à niveau vers Presto 0.125
- Mise à niveau d'Oozie vers la version 4.2.0
- Mise à niveau de Zeppelin vers la version 0.5.5

- Mise à niveau de la AWS SDK for Java version 1.10.27

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème en raison EMRFS CLI duquel le nom de la table de métadonnées par défaut n'était pas utilisé.
- Correction d'un problème rencontré lors de l'utilisation de tables ORC sauvegardées par -back dans Amazon S3.
- Résolution d'un problème rencontré avec une incompatibilité de version Python dans la configuration de Spark.
- Correction d'un problème lorsque l'état d'un YARN nœud n'était pas signalé en raison de DNS problèmes liés aux clusters d'unVPC.
- Correction d'un problème rencontré lors de la YARN mise hors service des nœuds, qui entraînait le blocage des applications ou l'impossibilité de planifier de nouvelles applications.
- Correction d'un problème rencontré lorsque les clusters se terminaient avec le statut TIMED _ OUT _STARTING.
- Correction d'un problème rencontré lors de l'inclusion de la dépendance EMRFS Scala dans d'autres versions. La dépendance Scala a été supprimée.

Versions des composants 4.2.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé myapp-component avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de 2.2-amzn-2.

Composant	Version	Description
emr-ddb	3.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.1.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
emr-s3-dist-cp	2.0.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.2.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
ganglia-monitor	3.6.0	Agent Ganglia intégré pour les applications de l'écosystème Hadoop avec agent de surveillance Ganglia.
ganglia-metadata-collector	3.6.0	Collecteur de métadonnées Ganglia pour agréger les métriques des agents de surveillance Ganglia.
ganglia-web	3.5.10	Application web pour afficher les métriques collectées par le collecteur de métadonnées Ganglia.

Composant	Version	Description
hadoop-client	2.6.0-amzn-2	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.6.0-amzn-2	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.6.0-amzn-2	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.6.0-amzn-2	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-https-server	2.6.0-amzn-2	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.6.0-amzn-2	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.6.0-amzn-2	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.6.0-amzn-2	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.6.0-amzn-2	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.

Composant	Version	Description
hive-client	1.0.0-amzn-1	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-5	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.11.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5	Mon serveur SQL de base de données.
oozie-client	4.2.0	Client de ligne de commande Oozie.
oozie-server	4.2.0	Service pour l'acceptation des demandes de flux de travail Oozie.
presto-coordinator	0.125	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.125	Service pour exécuter les éléments d'une requête.

Composant	Version	Description
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.5.2	Clients de ligne de commande Spark.
spark-history-server	1.5.2	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.5.2	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.5.2	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
webserver	2,4	HTTPserveur Apache.
zeppelin-server	0.5.5-incubating-amzn-0	Portable basé sur le web qui permet l'analyse de données interactives.

Classifications de configuration 4.2.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.2.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier <code>capacity-scheduler.xml</code> de Hadoop.

Classifications	Description
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.

Classifications	Description
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifiez les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.
presto-connector-hive	Modifiez les valeurs dans le fichier hive.properties de Presto.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
spark-metrics	Modifiez les valeurs dans le fichier metrics.properties de Spark.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Amazon EMR version 4.1.0

Versions de l'application 4.1.0

Les applications suivantes sont prises en charge dans cette version : [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) et [Zeppelin-Sandbox](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)

- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)
- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Kit AWS SDK pour Java	1,1,27	1,1,27	Non suivie	Non suivie
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
JupyterHub	-	-	-	-
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,5	0,5,5	0,6,0-SNAPSHOT	-
ZooKeeper	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
ZooKeeper-Sandbox	-	-	-	-

notes de publication de la version 4.1.0

Versions des composants 4.1.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme `CommunityVersion-amzn-EmrVersion`. `EmrVersion` commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
emr-ddb	3.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
emr-goodies	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
emr-kinesis	3.1.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.

Composant	Version	Description
emr-s3-dist-cp	2.0.0	Application de copie distribuée optimisée pour Amazon S3.
emrfs	2.1.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.
hadoop-client	2.6.0-amzn-1	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.6.0-amzn-1	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-library	2.6.0-amzn-1	HDFSclient et bibliothèque en ligne de commande
hadoop-hdfs-namenode	2.6.0-amzn-1	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httpfs-server	2.6.0-amzn-1	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-kms-server	2.6.0-amzn-1	Serveur de gestion des clés cryptographiques basé sur Hadoop. KeyProvider API
hadoop-mapred	2.6.0-amzn-1	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.6.0-amzn-1	YARNservice de gestion des conteneurs sur un nœud individuel.

Composant	Version	Description
hadoop-yarn-resourcemanager	2.6.0-amzn-1	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hive-client	1.0.0-amzn-1	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-1	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.
hive-server	1.0.0-amzn-1	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
hue-server	3.7.1-amzn-4	Application web pour l'analyse des données à l'aide d'applications de l'écosystème Hadoop
mahout-client	0.11.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5	Mon serveur SQL de base de données.
oozie-client	4.0.1	Client de ligne de commande Oozie.
oozie-server	4.0.1	Service pour l'acceptation des demandes de flux de travail Oozie.

Composant	Version	Description
presto-coordinator	0.119	Service pour accepter les requêtes et gérer l'exécution des requêtes des composants presto-worker.
presto-worker	0.119	Service pour exécuter les éléments d'une requête.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.5.0	Clients de ligne de commande Spark.
spark-history-server	1.5.0	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.5.0	Moteur d'exécution en mémoire pour. YARN
spark-yarn-slave	1.5.0	Bibliothèques Apache Spark nécessaires YARN aux esclaves.
zeppelin-server	0.6.0-incubation- SNAPSHOT	Portable basé sur le web qui permet l'analyse de données interactives.

4.1.0 classifications de configuration

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.1.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hdfs-encryption-zones	Configurez les zones de HDFS chiffrement.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
hue-ini	Modifiez les valeurs dans le fichier ini de Hue
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.

Classifications	Description
hadoop-kms-acls	Modifiez les valeurs dans le fichier kms-acls.xml de Hadoop.
hadoop-kms-env	Modifiez les valeurs dans l'environnement Hadoop. KMS
hadoop-kms-log4j	Modifiez les valeurs dans le fichier kms-log4j.properties de Hadoop.
hadoop-kms-site	Modifiez les valeurs dans le fichier kms-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
oozie-env	Modifiez les valeurs dans l'environnement d'Oozie.
oozie-log4j	Modifier les valeurs dans le fichier oozie-log4j.properties d'Oozie.
oozie-site	Modifiez les valeurs dans le fichier oozie-site.xml d'Oozie.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
presto-log	Modifiez les valeurs dans le fichier log.properties de Presto.
presto-config	Modifiez les valeurs dans le fichier config.properties de Presto.

Classifications	Description
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
yarn-env	Modifiez les valeurs dans l'YARN environnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.
zeppelin-env	Modifiez les valeurs dans l'environnement Zeppelin.

Amazon EMR version 4.0.0

Versions de l'application 4.0.0

Les applications suivantes sont prises en charge dans cette version : [Hadoop](#), [Hive](#), [Mahout](#), [Pig](#) et [Spark](#).

Le tableau ci-dessous répertorie les versions d'application disponibles dans cette version d'Amazon EMR et les versions d'application des trois EMR versions précédentes d'Amazon (le cas échéant).

Pour obtenir un historique complet des versions des applications pour chaque version d'AmazonEMR, consultez les rubriques suivantes :

- [Versions des applications dans les versions EMR 7.x d'Amazon](#)
- [Versions des applications dans les versions 6.x d'Amazon EMR](#)
- [Versions des applications dans les versions 5.x d'Amazon EMR](#)

- [Versions des applications dans les versions 4.x d'Amazon EMR](#)

Informations sur la version de l'application

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Kit AWS SDK pour Java	1,1,27	1,1,27	Non suivie	Non suivie
Python	Non suivie	Non suivie	Non suivie	Non suivie
Scala	Non suivie	Non suivie	Non suivie	Non suivie
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.6.0	-	-
HBase	-	-	-	-
HCatalog	-	-	-	-
Hadoop	2.7.1	2.6.0	2.6.0	2.6.0
Hive	1.0.0	1.0.0	1.0.0	1.0.0
Hudi	-	-	-	-
Hue	3.7.1	3.7.1	3.7.1	-
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-
JupyterHub	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Livy	-	-	-	-
MXNet	-	-	-	-
Mahout	0.11.0	0.11.0	0.11.0	0.10.0
Oozie	-	-	-	-
Oozie-Sandbox	4.2.0	4.2.0	4.0.1	-
Phoenix	-	-	-	-
Pig	0,14,0	0,14,0	0,14,0	0,14,0
Presto	-	-	-	-
Presto-Sandbox	0.130	0.125	0.119	-
Spark	1.6.0	1.5.2	1.5.0	1.4.1
Sqoop	-	-	-	-
Sqoop-Sandbox	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0,5,5	0,5,5	0,6,0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

notes de publication de la version 4.0.0

Versions des composants 4.0.0

Les composants qu'Amazon EMR installe avec cette version sont répertoriés ci-dessous. Certains sont installés dans le cadre de packages d'application de Big Data. D'autres sont propres à Amazon EMR et installés pour les processus et fonctionnalités du système. Ceux-ci commencent généralement par `emr` ou `aws`. Les packages d'applications Big Data de la dernière EMR version d'Amazon sont généralement les dernières versions disponibles dans la communauté. Nous mettons les publications communautaires à disposition sur Amazon EMR le plus rapidement possible.

Certains composants d'Amazon EMR diffèrent des versions communautaires. Ces composants ont une étiquette de version sous la forme *CommunityVersion*-amzn-*EmrVersion*. *EmrVersion* commence à 0. Par exemple, si un composant communautaire open source nommé `myapp-component` avec la version 2.2 a été modifié trois fois pour être inclus dans différentes EMR versions d'Amazon, sa version de publication est répertoriée sous le nom de `2.2-amzn-2`.

Composant	Version	Description
<code>emr-ddb</code>	3.0.0	Connecteur Amazon DynamoDB pour les applications de l'écosystème Hadoop.
<code>emr-goodies</code>	2.0.0	Bibliothèques proposant plus de commodités pour l'écosystème Hadoop.
<code>emr-kinesis</code>	3.0.0	Connecteur Amazon Kinesis pour les applications de l'écosystème Hadoop.
<code>emr-s3-dist-cp</code>	2.0.0	Application de copie distribuée optimisée pour Amazon S3.
<code>emrfs</code>	2.0.0	Connecteur Amazon S3 pour les applications de l'écosystème Hadoop.

Composant	Version	Description
hadoop-client	2.6.0-amzn-0	Clients de ligne de commande Hadoop tels que « hdfs », « hadoop » ou « yarn ».
hadoop-hdfs-datanode	2.6.0-amzn-0	HDFSservice au niveau du nœud pour le stockage de blocs.
hadoop-hdfs-namenode	2.6.0-amzn-0	HDFSservice de suivi des noms de fichiers et de l'emplacement des blocs.
hadoop-httfs-server	2.6.0-amzn-0	HTTPpoint de terminaison pour HDFS les opérations.
hadoop-mapred	2.6.0-amzn-0	MapReduce bibliothèques de moteurs d'exécution pour exécuter une MapReduce application.
hadoop-yarn-nodemanager	2.6.0-amzn-0	YARNservice de gestion des conteneurs sur un nœud individuel.
hadoop-yarn-resourcemanager	2.6.0-amzn-0	YARNservice d'allocation et de gestion des ressources du cluster et des applications distribuées.
hive-client	1.0.0-amzn-0	Client de ligne de commande Hive.
hive-metastore-server	1.0.0-amzn-0	Service d'accès au métastore Hive, un référentiel sémantique et stockant des métadonnées pour SQL les opérations sur Hadoop.

Composant	Version	Description
hive-server	1.0.0-amzn-0	Service pour l'acceptation de requêtes Hive en tant que requêtes web.
mahout-client	0.10.0	Bibliothèque pour la technologie Machine Learning.
mysql-server	5.5	Mon serveur SQL de base de données.
pig-client	0.14.0-amzn-0	Client de ligne de commande Pig.
spark-client	1.4.1	Clients de ligne de commande Spark.
spark-history-server	1.4.1	Interface web pour afficher des événements enregistrés pour la durée de vie d'une application Spark terminée.
spark-on-yarn	1.4.1	Moteur d'exécution en mémoire pour YARN
spark-yarn-slave	1.4.1	Bibliothèques Apache Spark nécessaires YARN aux esclaves.

Classifications de configuration 4.0.0

Les classifications de configuration vous permettent de personnaliser les applications. Ils correspondent souvent à un XML fichier de configuration de l'application, tel que `hive-site.xml`. Pour plus d'informations, consultez [Configuration des applications](#).

Classifications emr-4.0.0

Classifications	Description
capacity-scheduler	Modifiez les valeurs dans le fichier capacity-scheduler.xml de Hadoop.
core-site	Modifiez les valeurs dans le fichier core-site.xml de Hadoop.
emrfs-site	Modifiez EMRFS les paramètres.
hadoop-env	Modifiez les valeurs dans l'environnement Hadoop pour tous les composants Hadoop.
hadoop-log4j	Modifiez les valeurs dans le fichier log4j.properties de Hadoop.
hdfs-site	Modifiez les valeurs dans HDFS le fichier hdfs-site.xml.
hive-env	Modifiez les valeurs dans l'environnement Hive.
hive-exec-log4j	Modifiez les valeurs dans le fichier hive-exec-log4j.properties de Hive.
hive-log4j	Modifiez les valeurs dans le fichier hive-log4j.properties de Hive.
hive-site	Modifiez les valeurs dans le fichier hive-site.xml de Hive
httpfs-env	Modifiez les valeurs dans l'HTTPFS environnement.
httpfs-site	Modifiez les valeurs dans le fichier httpfs-site.xml de Hadoop.
mapred-env	Modifiez les valeurs dans l'environnement de l' MapReduce application.

Classifications	Description
mapred-site	Modifiez les valeurs dans le fichier mapred-site.xml de l' MapReduce application.
pig-properties	Modifiez les valeurs dans le fichier pig.properties de Pig.
pig-log4j	Modifiez les valeurs dans le fichier log4j.properties de Pig.
spark	Paramètres EMR sélectionnés par Amazon pour Apache Spark.
spark-defaults	Modifiez les valeurs dans le fichier spark-defaults.conf de Spark.
spark-env	Modifiez les valeurs dans l'environnement Spark.
spark-log4j	Modifiez les valeurs dans le fichier log4j.properties de Spark.
yarn-env	Modifiez les valeurs dans l'YARNenvironnement.
yarn-site	Modifiez les valeurs dans YARN le fichier yarn-site.xml.

Versions Amazon EMR 2.x et AMI 3.x

Note

AWS met à jour la TLS configuration de tous les AWS API points de terminaison vers une version minimale de TLS 1.2. EMR Les versions 3.10 et antérieures d'Amazon ne prennent en charge que les connexions TLS 1.0/1.1. Après le 4 décembre 2023, vous ne pourrez plus créer de clusters avec Amazon EMR 3.10 et versions antérieures.

Si vous utilisez Amazon EMR 3.10 ou une version antérieure, nous vous recommandons de tester et de migrer immédiatement vos charges de travail vers la dernière version d'AmazonEMR. Pour plus d'informations, consultez le [Blog sur la sécurité d'AWS](#).

Les versions EMR 2.x et 3.x d'Amazon, appelées AMI versions, sont mises à disposition pour les solutions préexistantes qui en ont besoin pour des raisons de compatibilité. Il est déconseillé de créer de nouveaux clusters ou de nouvelles solutions avec ces versions. Elles ne sont pas dotées des fonctions présentes dans les versions plus récentes et contiennent des packages d'application obsolètes.

Nous vous recommandons de créer des solutions à l'aide de la EMR version la plus récente d'Amazon.

L'étendue des différences entre les versions des séries 2.x et 3.x et les versions récentes d'Amazon EMR est significative. Ces différences s'appliquent à différents domaines, comme la création et la configuration d'un cluster, ou encore les ports et la structure de répertoire des applications sur le cluster.

Cette section tente de couvrir les différences les plus importantes pour AmazonEMR, ainsi que les différences spécifiques en matière de configuration et de gestion des applications. Elle n'est pas exhaustive. Si vous créez et utilisez des clusters dans les versions 2.x ou 3.x, il se peut que vous rencontriez des différences non répertoriées dans cette section.

Rubriques

- [Création d'un cluster avec des AMI versions antérieures d'Amazon EMR](#)
- [Installation d'applications avec des AMI versions antérieures d'Amazon EMR](#)
- [Personnalisation de la configuration des clusters et des applications avec AMI les versions antérieures d'Amazon EMR](#)
- [Spécificités de l'application Hive pour les AMI versions antérieures d'Amazon EMR](#)
- [HBasespécificités de l'application pour AMI les versions antérieures d'Amazon EMR](#)
- [Spécificités de l'application Pig pour AMI les versions antérieures d'Amazon EMR](#)
- [Spécificités de l'application Spark avec AMI les versions antérieures d'Amazon EMR](#)
- [Différences entre l'DistCp utilitaire S3 et AMI les versions antérieures d'Amazon EMR](#)

Création d'un cluster avec des AMI versions antérieures d'Amazon EMR

Les versions EMR 2.x et 3.x d'Amazon sont référencées par AMI version. Avec les EMR versions 4.0.0 et ultérieures d'Amazon, les versions sont référencées par version, à l'aide d'une étiquette de version telle que `emr-5.11.0`. Ce changement est particulièrement visible lorsque vous créez un cluster à l'aide du AWS CLI ou par programmation.

Lorsque vous utilisez le AWS CLI pour créer un cluster à l'aide d'une AMI version finale, utilisez l'option `--ami-version`, par exemple, `--ami-version 3.11.0`. De nombreuses options, fonctionnalités et applications introduites dans Amazon EMR 4.0.0 et versions ultérieures ne sont pas disponibles lorsque vous spécifiez un `--ami-version`. Pour plus d'informations, consultez [create-cluster](#) dans la Référence des commandes de la AWS CLI .

L'exemple de AWS CLI commande suivant lance un cluster à l'aide d'une AMI version.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.11.0 \  
--applications Name=Hue Name=Hive Name=Pig \  
--use-default-roles --ec2-attributes KeyName=myKey \  
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,\  
InstanceType=m3.xlarge InstanceGroupType=CORE,InstanceCount=2,\  
InstanceType=m3.xlarge --bootstrap-actions Path=s3://elasticmapreduce/bootstrap-  
actions/configure-hadoop,\  
Name="Configuring infinite JVM reuse",Args=["-m", "mapred.job.reuse.jvm.num.tasks=-1"]
```

Du point de vue de la programmation, toutes les EMR versions d'Amazon utilisent l'opération `RunJobFlowRequest` décrite dans le EMR API pour créer des clusters. L'exemple de code Java suivant crée un cluster à l'aide de la AMI version 3.11.0.

```
RunJobFlowRequest request = new RunJobFlowRequest()  
    .withName("AmiVersion Cluster")  
    .withAmiVersion("3.11.0")  
    .withInstances(new JobFlowInstancesConfig()  
        .withEc2KeyName("myKeyPair"))
```



```
.withInstanceCount(1)
.withKeepJobFlowAliveWhenNoSteps(true)
.withMasterInstanceType("m3.xlarge")
.withSlaveInstanceType("m3.xlarge");
```

L'appel `RunJobFlowRequest` suivant utilise quant à lui une étiquette de version :

```
RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("ReleaseLabel Cluster")
    .withReleaseLabel("emr-7.2.0")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyPair")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge");
```

Configuration de la taille du cluster

Lorsque votre cluster s'exécute, Hadoop détermine le nombre de tâches de mappeur et de réducteur nécessaires pour traiter les données. Les clusters les plus grands doivent avoir plus de tâches pour une meilleure utilisation des ressources et une plus courte durée de traitement. Généralement, la taille d'un EMR cluster reste la même pendant toute la durée du cluster ; vous définissez le nombre de tâches lorsque vous créez le cluster. Lorsque vous redimensionnez un cluster en cours d'exécution, vous pouvez modifier le traitement au cours de l'exécution du cluster. Par conséquent, au lieu d'utiliser un nombre fixe de tâches, vous pouvez modifier le nombre de tâches au cours de la vie du cluster. Il existe deux options de configuration qui vous aideront à définir le nombre idéal de tâches :

- `mapred.map.tasksperslot`
- `mapred.reduce.tasksperslot`

Vous pouvez définir ces deux options dans le fichier `mapred-conf.xml`. Lorsque vous soumettez un travail au cluster, le client du travail vérifie le nombre total actuel d'emplacements de mappage et de réduction disponibles sur l'ensemble du cluster. Le client du travail utilise alors les équations suivantes pour définir le nombre de tâches :

- `mapred.map.tasks = mapred.map.tasksperslot * emplacements de mappage dans le cluster`

- `mapred.reduce.tasks = mapred.reduce.tasksperslot * emplacements de réduction dans le cluster`

Le client du travail lit uniquement le paramètre `tasksperslot` si le nombre de tâches n'est pas configuré. Vous pouvez remplacer le nombre de tâches à tout moment, pour tous les clusters via une action d'amorçage ou individuellement, travail par travail, en ajoutant une étape pour modifier la configuration.

Amazon EMR résiste aux défaillances des nœuds de tâches et poursuit l'exécution du cluster même si un nœud de tâche devient indisponible. Amazon EMR met automatiquement en service des nœuds de tâches supplémentaires pour remplacer ceux qui échouent.

Vous pouvez avoir un nombre différent de nœuds de tâches pour chaque étape du cluster. Vous pouvez également ajouter une étape à un cluster en cours d'exécution pour modifier le nombre de nœuds de tâches. Étant donné que toutes les étapes sont garanties de s'exécuter de manière séquentielle par défaut, vous pouvez spécifier le nombre de nœuds de tâches en cours d'exécution pour n'importe quelle étape.

Installation d'applications avec des AMI versions antérieures d'Amazon EMR

Lors de l'utilisation d'une AMI version, les applications sont installées de différentes manières, notamment en utilisant le `NewSupportedProducts` paramètre de l'[RunJobFlow](#) action, en utilisant les actions d'amorçage et en utilisant l'action [Step](#).

Personnalisation de la configuration des clusters et des applications avec AMI les versions antérieures d'Amazon EMR

La EMR version 4.0.0 d'Amazon a introduit une méthode simplifiée de configuration des applications à l'aide de classifications de configuration. Pour de plus amples informations, veuillez consulter [Configuration des applications](#). Lorsque vous utilisez une AMI version, vous configurez les applications à l'aide d'actions de démarrage ainsi que des arguments que vous transmettez. Par exemple, les actions `configure-hadoop` et `configure-daemons bootstrap` définissent des propriétés YARN spécifiques à Hadoop et à l'environnement, telles que `--namenode-heap-size`. Dans les versions plus récentes, ces propriétés sont configurées à l'aide des classifications de configuration `hadoop-env` et `yarn-env`. Pour les actions de bootstrap qui exécutent des configurations courantes, consultez le [emr-bootstrap-actions référentiel sur Github](#).

Les tableaux suivants associent les actions de bootstrap aux classifications de configuration des EMR versions les plus récentes d'Amazon.

Hadoop

Nom du fichier d'application concerné	AMlversion bootstrap action	Classification de configuration
core-site.xml	configure-hadoop -c	core-site
log4j.properties	configure-hadoop -l	hadoop-log4j
hdfs-site.xml	configure-hadoop -s	hdfs-site
N/A	s/o	hdfs-encryption-zones
mapred-site.xml	configure-hadoop -m	mapred-site
yarn-site.xml	configure-hadoop -y	yarn-site
httpfs-site.xml	configure-hadoop -t	httpfs-site
capacity-scheduler.xml	configure-hadoop -z	capacity-scheduler
yarn-env.sh	configure-daemons --resourcemanager-opts	yarn-env

Hive

Nom du fichier d'application concerné	AMlversion bootstrap action	Classification de configuration
hive-env.sh	N/A	hive-env
hive-site.xml	hive-script --install -hive-site \${MY_HIVE_SITE_FILE}	hive-site

Nom du fichier d'application concerné	AMIversion bootstrap action	Classification de configuration
hive-exec-log4j.properties	N/A	hive-exec-log4j
hive-log4j.properties	s/o	hive-log4j

EMRFS

Nom du fichier d'application concerné	AMIversion bootstrap action	Classification de configuration
emrfs-site.xml	configure-hadoop -e	emrfs-site
N/A	s3get -s s3://customer-provider.jar -d /usr/share/aws/emr/auxlib/	emrfs-site (avec le nouveau paramètre fs.s3.cse.encryptionMaterialsProvider.uri)

Pour obtenir une liste de toutes les classifications, consultez [Configuration des applications](#).

Variables d'environnement d'application

Lors de l'utilisation d'une AMI version, un `hadoop-user-env.sh` script est utilisé avec l'action `configure-daemons bootstrap` pour configurer l'environnement Hadoop. Les actions suivantes sont incluses dans le script :

```
#!/bin/bash
export HADOOP_USER_CLASSPATH_FIRST=true;
echo "HADOOP_CLASSPATH=/path/to/my.jar" >> /home/hadoop/conf/hadoop-user-env.sh
```

Dans Amazon EMR version 4.x, vous devez procéder de même en utilisant la classification `hadoop-env` de configuration, comme indiqué dans l'exemple suivant :

```
[
```

```

    {
      "Classification":"hadoop-env",
      "Properties":{

    },
    "Configurations":[
      {
        "Classification":"export",
        "Properties":{
          "HADOOP_USER_CLASSPATH_FIRST":"true",
          "HADOOP_CLASSPATH":"/path/to/my.jar"
        }
      }
    ]
  }
]

```

Autre exemple : l'utilisation de `configure-daemons` et la transmission de `--namenode-heap-size=2048` et `--namenode-opts=-XX:GCTimeRatio=19` équivalent aux classifications de configuration suivantes.

```

[
  {
    "Classification":"hadoop-env",
    "Properties":{

  },
  "Configurations":[
    {
      "Classification":"export",
      "Properties":{
        "HADOOP_DATANODE_HEAPSIZE": "2048",
        "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
      }
    }
  ]
}
]

```

D'autres variables de l'environnement d'application ne sont plus définies dans `/home/hadoop/.bashrc`. Au lieu de cela, elles sont principalement définies dans les fichiers `/etc/default` par composant ou par application, par exemple, `/etc/default/hadoop`. Les scripts

wrapper `/usr/bin/` installés par l'application RPMs peuvent également définir des variables d'environnement supplémentaires avant d'impliquer le script `bin` lui-même.


Ports de service

Lors de l'utilisation d'une AMI version, certains services utilisent des ports personnalisés.

Modifications des paramètres de port

Paramètre	AMI version 3.x	Open source par défaut
<code>fs.default.name</code>	<code>hdfs://IP:9000</code> <code>emrDetermined</code>	par défaut (<code>hdfs://<i>emrDeterminedIP</i>:8020</code>)
<code>dfs.datanode.address</code>	<code>0.0.0.0:9200</code>	valeur par défaut (<code>0.0.0.0:50010</code>)
<code>dfs.datanode.http.address</code>	<code>0.0.0.0:9102</code>	valeur par défaut (<code>0.0.0.0:50075</code>)
<code>dfs.datanode.https.address</code>	<code>0.0.0.0:9402</code>	valeur par défaut (<code>0.0.0.0:50475</code>)
<code>dfs.datanode.ipc.address</code>	<code>0.0.0.0:9201</code>	valeur par défaut (<code>0.0.0.0:50020</code>)
<code>dfs.http.address</code>	<code>0.0.0.0:9101</code>	valeur par défaut (<code>0.0.0.0:50070</code>)
<code>dfs.https.address</code>	<code>0.0.0.0:9202</code>	valeur par défaut (<code>0.0.0.0:50470</code>)
<code>dfs.secondary.http.address</code>	<code>0.0.0.0:9104</code>	valeur par défaut (<code>0.0.0.0:50090</code>)
<code>yarn.nodemanager.address</code>	<code>0.0.0.0:9103</code>	valeur par défaut (<code>\${yarn.nodemanager.hostname}:0</code>)
<code>yarn.nodemanager.localizer.address</code>	<code>0.0.0.0:9033</code>	valeur par défaut (<code>\${yarn.nodemanager.hostname}:8040</code>)

Paramètre	AMlversion 3.x	Open source par défaut
yarn.nodemanager.webapp.address	0.0.0.0:9035	valeur par défaut (\${yarn.nodemanager.hostname}:8042)
yarn.resourcemanager.address	<i>emrDeterminedIP</i> :9022	valeur par défaut (\${yarn.resourcemanager.hostname}:8032)
yarn.resourcemanager.admin.address	<i>emrDeterminedIP</i> :9025	valeur par défaut (\${yarn.resourcemanager.hostname}:8033)
yarn.resourcemanager.resource-tracker.address	<i>emrDeterminedIP</i> :9023	valeur par défaut (\${yarn.resourcemanager.hostname}:8031)
yarn.resourcemanager.scheduler.address	<i>emrDeterminedIP</i> :9024	valeur par défaut (\${yarn.resourcemanager.hostname}:8030)
yarn.resourcemanager.webapp.address	0.0.0.0:9026	valeur par défaut (\${yarn.resourcemanager.hostname}:8088)
yarn.web-proxy.address	<i>emrDeterminedIP</i> :9046	valeur par défaut (no-value)
yarn.resourcemanager.hostname	0.0.0.0 (par défaut)	<i>emrDeterminedIP</i>

 Note

Le *emrDeterminedIP* est une adresse IP générée par AmazonEMR.

Users

Lors de l'utilisation d'une AMI version, l'utilisateur hadoop exécute tous les processus et possède tous les fichiers. Dans les EMR versions 4.0.0 et ultérieures d'Amazon, les utilisateurs existent au niveau de l'application et du composant.

Séquence d'installation, artefacts installés et emplacement des fichiers journaux

Lors de l'utilisation d'une AMI version, les artefacts de l'application et leurs répertoires de configuration sont installés dans le `/home/hadoop/application` répertoire. Par exemple, si vous avez installé Hive, le répertoire est `/home/hadoop/hive`. Dans les EMR versions 4.0.0 et ultérieures d'Amazon, les artefacts de l'application sont installés dans le `/usr/lib/application` répertoire. Lors de l'utilisation d'une AMI version, les fichiers journaux se trouvent à différents endroits. Le tableau ci-dessous répertorie ces emplacements.

Modification des emplacements des journaux sur Amazon S3

Démon ou application	Emplacement du répertoire
instance-state	noeud/ <i>instance-id</i> /état de l'instance/
hadoop-hdfs-namenode	démons/ <i>instance-id</i> /hadoop-hadoop-namenode.journal
hadoop-hdfs-datanode	démons/ <i>instance-id</i> /hadoop-hadoop-datanode.journal
fil de hadoop () ResourceManager	démons/ <i>instance-id</i> /yarn-hadoop-resource-manager
hadoop-yarn (serveur Proxy)	démons/ <i>instance-id</i> /yarn-hadoop-proxy-server
mapred-historyserver	démons/ <i>instance-id</i> /
httpfs	démons/ <i>instance-id</i> /httpfs.log
hive-server	noeud/ <i>instance-id</i> /hive-server/hive-server.log
hive-metastore	noeud/ <i>instance-id</i> /apps/hive.log

Démon ou application	Emplacement du répertoire
Ruche CLI	noeud/ <i>instance-id</i> /apps/hive.log
YARN journaux des utilisateurs des applications et journaux des conteneurs	task-attempts/
Mahout	N/A
Pig	N/A
spark-historyserver	N/A
Fichiers d'historique des tâches mapreduce	jobs/

Exécuteur de commandes

Lorsque vous utilisez une AMI version, de nombreux scripts ou programmes, par exemple /home/hadoop/contrib/streaming/hadoop-streaming.jar, ne sont pas placés dans l'environnement du chemin de connexion du shell. Vous devez donc spécifier le chemin complet lorsque vous utilisez un fichier jar tel que command-runner.jar ou script-runner.jar pour exécuter les scripts. Le command-runner.jar est situé sur le AMI, il n'est donc pas nécessaire d'en connaître le complet URI comme c'était le cas pour script-runner.jar.

Facteur de réplication

Le facteur de réplication vous permet de configurer à quel moment démarrer un Hadoop JVM. Vous pouvez démarrer un nouvel Hadoop JVM pour chaque tâche, ce qui permet une meilleure isolation des tâches, ou vous pouvez partager les tâches JVMs entre les tâches, ce qui permet de réduire les frais de structure. Si vous traitez de nombreux petits fichiers, il est judicieux de les réutiliser JVM plusieurs fois afin d'amortir le coût de démarrage. Toutefois, si chaque tâche prend du temps ou traite une grande quantité de données, vous pouvez choisir de ne pas les réutiliser JVM afin de libérer toute la mémoire pour les tâches suivantes. Lorsque vous utilisez une AMI version, vous pouvez personnaliser le facteur de réplication à l'aide de l'action configure-hadoop bootstrap pour définir la mapred.job.reuse.jvm.num.tasks propriété.

L'exemple suivant montre comment définir le facteur de JVM réutilisation pour une JVM réutilisation infinie.

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).




```
aws emr create-cluster --name "Test cluster" --ami-version 3.11.0 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m3.xlarge \
InstanceGroupType=CORE,InstanceCount=2,InstanceType=m3.xlarge \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hadoop,\
Name="Configuring infinite JVM reuse",Args=["-m", "mapred.job.reuse.jvm.num.tasks=-1"]
```

Spécificités de l'application Hive pour les AMI versions antérieures d'Amazon EMR

Les fichiers journaux

Avec EMR AMI les versions 2.x et 3.x d'Amazon, les journaux Hive sont enregistrés dans. /mnt/var/log/apps/ La version de Hive que vous exécutez détermine le nom du fichier journal, comme illustré dans le tableau suivant. Cela permet de prendre en charge les versions simultanées de Hive.

Version de Hive	Nom du fichier journal
0.13.1	hive.log
	Note À partir de cette version, Amazon EMR utilise un nom de fichier non versionné, .hive.log Les versions mineures partagent le même emplacement du journal que la version majeure.
0.11.0	hive_0110.log

Version de Hive	Nom du fichier journal
	<p> Note</p> <p>Les versions mineures de Hive 0.11.0, par exemple 0.11.0.1, partagent le même emplacement du fichier journal que Hive 0.11.0.</p>
0.8.1	<p>hive_081.log</p> <p> Note</p> <p>Les versions mineures de Hive 0.8.1, par exemple Hive 0.8.1.1, partagent le même emplacement du fichier journal que Hive 0.8.1.</p>
0.7.1	<p>hive_07_1.log</p> <p> Note</p> <p>Les versions mineures de Hive 0.7.1, par exemple Hive 0.7.1.3 et Hive 0.7.1.4, partagent le même emplacement du fichier journal que Hive 0.7.1.</p>
0.7	hive_07.log
0.5	hive_05.log
0.4	hive.log

Fonctionnalité de format d'entrée avec fractionnement

Pour implémenter la fonctionnalité de saisie partagée à l'aide des versions de Hive antérieures à 0.13.1 (EMRAMIversions Amazon antérieures à 3.11.0), utilisez ce qui suit :

```
hive> set hive.input.format=org.apache.hadoop.hive.q1.io.HiveCombineSplitsInputFormat;
```

```
hive> set mapred.min.split.size=100000000;
```

Cette fonctionnalité est devenue obsolète avec Hive 0.13.1. Pour bénéficier de la même fonctionnalité de format d'entrée fractionné dans Amazon EMR AMI version 3.11.0, utilisez ce qui suit :

```
set hive.hadoop.supports.plittable.combineinputformat=true;
```

Ports de service Thrift

Thrift est un RPC framework qui définit un format de sérialisation binaire compact utilisé pour conserver les structures de données en vue d'une analyse ultérieure. Généralement, Hive configure le serveur pour qu'il fonctionne sur les ports suivants.

Version de Hive	Numéro de port
Hive 0.13.1	10 000
Hive 0.11.0	10004
Hive 0.8.1	10003
Hive 0.7.1	10002
Hive 0.7	10001
Hive 0.5	10 000

Pour plus d'informations sur les services thrift, consultez <http://wiki.apache.org/thrift/>.

Utilisation de Hive pour récupérer des partitions

Amazon EMR inclut une instruction dans le langage de requête Hive qui permet de récupérer les partitions d'une table à partir des données de table situées dans Amazon S3. L'exemple suivant en est une illustration.

```
CREATE EXTERNAL TABLE (json string) raw_impression
```

```
PARTITIONED BY (dt string)
LOCATION 's3://elastic-mapreduce/samples/hive-ads/tables/impressions';
ALTER TABLE logs RECOVER PARTITIONS;
```

Les données et les répertoires de la partition doit être à l'emplacement spécifié dans la définition de table et doivent être nommées selon la convention Hive : par exemple, dt=2009-01-01.

Note

Après Hive 0.13.1, cette fonctionnalité est prise en charge en mode natif à l'aide de `msck repair table` et par conséquent la prise en charge de `recover partitions` n'est plus assurée. Pour plus d'informations, consultez <https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL>.

Transmission d'une variable Hive à un script

Pour passer une variable dans une étape Hive à l'aide de AWS CLI, tapez la commande suivante, remplacez *myKey* avec le nom de votre paire de EC2 clés, et remplacez *mybucket* avec le nom de votre compartiment. Dans cet exemple, SAMPLE est une valeur de variable précédée par le commutateur `-d`. Cette variable est définie dans le script Hive comme suit : `${SAMPLE}`.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.9 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type m3.xlarge --instance-count 3 \
--steps Type=Hive,Name="Hive Program",ActionOnFailure=CONTINUE,\
Args=[-f,s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q,-d,\
INPUT=s3://elasticmapreduce/samples/hive-ads/tables,-d,OUTPUT=s3://mybucket/hive-ads/\
output/,\
-d,SAMPLE=s3://elasticmapreduce/samples/hive-ads/]
```

Spécification d'un emplacement de metastore externe

La procédure suivante vous montre comment remplacer les valeurs de configuration par défaut pour l'emplacement du metastore Hive et démarrer un cluster en utilisant l'emplacement du metastore reconfiguré.

Pour créer un métastore situé en dehors du cluster EMR

1. Créez une base de données My SQL ou Aurora à l'aide d'AmazonRDS.

Pour plus d'informations sur la création d'une RDS base de données Amazon, consultez [Getting started with Amazon RDS](#).

2. Modifiez vos groupes de sécurité pour autoriser JDBC les connexions entre votre base de données et le groupe de sécurité ElasticMapReduce-Master.

Pour plus d'informations sur la façon de modifier vos groupes de sécurité pour l'accès, consultez la section [Groupes RDS de sécurité Amazon](#) dans le guide de RDS l'utilisateur Amazon.

3. Définissez les valeurs JDBC de configuration dans `hive-site.xml` :
 - a. Créez un fichier de configuration `hive-site.xml` contenant les éléments suivants :

```
<configuration>
  <property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:mariadb://hostname:3306/hive?createDatabaseIfNotExist=true</
value>
    <description>JDBC connect string for a JDBC metastore</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionUserName</name>
    <value>hive</value>
    <description>Username to use against metastore database</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionPassword</name>
    <value>password</value>
    <description>Password to use against metastore database</description>
  </property>
</configuration>
```

hostname est l'adresse DNS de l'instance Amazon RDS qui exécute la base de données. *username* and *password* sont les informations d'identification de votre base de données. Pour plus d'informations sur la connexion aux instances de base de données My SQL et Aurora, consultez les sections [Connexion à une instance de base de données exécutant le moteur de SQL base de données My](#) et [Connexion à un cluster de base de données Aurora](#) dans le guide de RDS l'utilisateur Amazon.

Les JDBC pilotes sont installés par AmazonEMR.

Note

La propriété `value` ne doit pas contenir d'espaces ni de retours chariot. Elle doit figurer entièrement sur une seule ligne.

- b. Enregistrez votre fichier `hive-site.xml` dans un emplacement sur Amazon S3, tel que `s3://mybucket/hive-site.xml`.
4. Créez un cluster, en spécifiant l'emplacement Amazon S3 du fichier `hive-site.xml` personnalisé.

L'exemple de commande suivant illustre une AWS CLI commande qui effectue cette opération.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.10 \
--applications Name=Hue Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type m3.xlarge --instance-count 3 \
--bootstrap-actions Name="Install Hive Site Configuration",\
Path="s3://region.elasticmapreduce/libs/hive/hive-script",\
Args=["--base-path","s3://elasticmapreduce/libs/hive","--install-hive-site",\
"--hive-site=s3://mybucket/hive-site.xml","--hive-versions","latest"]
```

Connect to Hive en utilisant JDBC

Pour vous connecter à Hive via JDBC, vous devez télécharger le JDBC pilote et installer un SQL client. L'exemple suivant montre comment utiliser SQL Workbench/J pour se connecter à Hive en utilisant. JDBC

Pour télécharger des JDBC pilotes

1. Téléchargez et extrayez les pilotes correspondant aux versions de Hive auxquelles vous souhaitez accéder. La version Hive varie en fonction de celle AMI que vous choisissez lorsque vous créez un EMR cluster Amazon.
 - [Pilotes Hive 0.13.1 JDBC : _1.0.4.1004.zip https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/ AmazonHive JDBC](https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/AmazonHiveJDBC)
 - Pilotes Hive 0.11.0 JDBC : <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.11.0>
 - Pilotes Hive 0.8.1 JDBC : <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.8.1>
2. Installez SQL Workbench/J. Pour plus d'informations, voir [Installation et démarrage de SQL Workbench/J dans le manuel d'utilisation du SQL manuel Workbench/J](#).
3. Créez un SSH tunnel vers le nœud principal du cluster. Le port de connexion diffère en fonction de la version de Hive. Des exemples de commandes sont fournis dans les tableaux ci-dessous pour les ssh utilisateurs de Linux et des TTY commandes Pu pour les utilisateurs de Windows.

SSHCommandes Linux

Version de Hive	Command
0.13.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10000:localhost:10000 hadoop@ <i>master-public-dns-name</i></code>
0.11.0	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10004:localhost:10004 hadoop@ <i>master-public-dns-name</i></code>
0.8.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10003:localhost:10003 hadoop@ <i>master-public-dns-name</i></code>
0.7.1	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10002:localhost:10002 hadoop@ <i>master-public-dns-name</i></code>

Version de Hive	Command
0.7	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10001:localhost:10001 hadoop@ <i>master-public-dns-name</i></code>
0.5	<code>ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10000:localhost:10000 hadoop@ <i>master-public-dns-name</i></code>

Paramètres du TTY tunnel Windows Pu

Version de Hive	Paramètres de tunnel
0.13.1	Port source : 10000 Destination : <i>master-public-dns-name</i> :10000
0.11.0	Port source : 10004 Destination : <i>master-public-dns-name</i> :10004
0.8.1	Port source : 10003 Destination : <i>master-public-dns-name</i> :10003

4. Ajoutez le JDBC pilote à SQL Workbench.
 - a. Dans la boîte de dialogue Select Connection Profile (Sélectionner le profil de connexion), sélectionnez Manage Drivers (Gérer les pilotes).
 - b. Sélectionnez l'icône Create a new entry (Créer une nouvelle saisie) (page blanche).
 - c. Dans le champ Name (Nom), saisissez **Hive JDBC**.
 - d. Pour Bibliothèque, cliquez sur l'icône Sélectionner le ou les JAR fichiers.
 - e. Sélectionnez JAR les fichiers comme indiqué dans le tableau suivant.

Version du pilote Hive	JARfichiers à ajouter
0.13.1	<pre>hive_metastore.jar hive_service.jar HiveJDBC3.jar libfb303-0.9.0.jar</pre>

Version du pilote Hive	JARfichiers à ajouter
	<pre>libthrift-0.9.0.jar log4j-1.2.14.jar ql.jar slf4j-api-1.5.8.jar slf4j-log4j12-1.5.8.jar TCLIServiceClient.jar</pre>
0.11.0	<pre>hadoop-core-1.0.3.jar hive-exec-0.11.0.jar hive-jdbc-0.11.0.jar hive-metastore-0.11.0.jar hive-service-0.11.0.jar libfb303-0.9.0.jar commons-logging-1.0.4.jar slf4j-api-1.6.1.jar</pre>
0.8.1	<pre>hadoop-core-0.20.205.jar hive-exec-0.8.1.jar hive-jdbc-0.8.1.jar hive-metastore-0.8.1.jar hive-service-0.8.1.jar libfb303-0.7.0.jar libthrift-0.7.0.jar log4j-1.2.15.jar slf4j-api-1.6.1.jar slf4j-log4j12-1.6.1.jar</pre>
0.7.1	<pre>hadoop-0.20-core.jar hive-exec-0.7.1.jar hive-jdbc-0.7.1.jar hive-metastore-0.7.1.jar hive-service-0.7.1.jar libfb303.jar commons-logging-1.0.4.jar slf4j-api-1.6.1.jar slf4j-log4j12-1.6.1.jar</pre>

Version du pilote Hive	JARfichiers à ajouter
0.7	<pre>hadoop-0.20-core.jar hive-exec-0.7.0.jar hive-jdbc-0.7.0.jar hive-metastore-0.7.0.jar hive-service-0.7.0.jar libfb303.jar commons-logging-1.0.4.jar slf4j-api-1.5.6.jar slf4j-log4j12-1.5.6.jar</pre>
0.5	<pre>hadoop-0.20-core.jar hive-exec-0.5.0.jar hive-jdbc-0.5.0.jar hive-metastore-0.5.0.jar hive-service-0.5.0.jar libfb303.jar log4j-1.2.15.jar commons-logging-1.0.4.jar</pre>

- f. Dans la boîte de dialogue Please select one driver (Veuillez sélectionner un pilote), sélectionnez un pilote en vous conformant au tableau suivant, puis cliquez sur OK.

Version de Hive	Nom de classe du pilote
0.13.1	<code>com.amazon.hive.jdbc3.HS2Driver</code>
0.11.0	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.8.1	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.7.1	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>

Version de Hive	Nom de classe du pilote
0.7	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0.5	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>

5. Lorsque vous revenez à la boîte de dialogue Sélectionner le profil de connexion, vérifiez que le champ Driver est défini sur Hive JDBC et saisissez la chaîne de JDBC connexion dans le URLchamp conformément au tableau suivant.

Version de Hive	JDBCchaîne de connexion
0.13.1	<code>jdbc:hive2://localhost:10000/default</code>
0.11.0	<code>jdbc:hive://localhost:10004/default</code>
0.8.1	<code>jdbc:hive://localhost:10003/default</code>

Si votre cluster utilise AMI la version 3.3.1 ou ultérieure, dans la boîte de dialogue Sélectionner le profil de connexion, saisissez **hadoop** le champ Nom d'utilisateur.

HBasespécificités de l'application pour AMI les versions antérieures d'Amazon EMR

HBaseVersions prises en charge

HBaseversion	AMIversion	AWS CLI paramètres de configuration	HBasedétails de la version
0.94.18	3.1.0 et ultérieures	<code>--ami-version 3.1</code>	<ul style="list-style-type: none"> • Correctifs de bogues et améliorations

HBaseversion	AMIversion	AWS CLI paramètres de configuration	HBasedétails de la version
		<pre>--ami-version 3.2 --ami-version 3.3 --applications Name=HBase</pre>	
0.94.7	3.0-3.0.4	<pre>--ami-version 3.0 --applications Name=HBase</pre>	
0.92	2.2 et ultérieures	<pre>--ami-version 2.2 or later --applications Name=HBase</pre>	

HBaseprérequis du cluster

Un cluster créé à l'aide des EMR AMI versions 2.x et 3.x d'Amazon doit répondre aux exigences suivantes pour HBase

- Le AWS CLI (facultatif) —Pour interagir à HBase l'aide de la ligne de commande, téléchargez et installez la dernière version du AWS CLI. Pour plus d'informations, consultez [Installation d' AWS Command Line Interface](#) dans le Guide de l'utilisateur AWS Command Line Interface .
- Au moins deux instances (facultatif) : le nœud principal du cluster exécute le serveur HBase principal et Zookeeper, et les nœuds de tâches exécutent les serveurs HBase régionaux. Pour de meilleures performances, les HBase clusters doivent s'exécuter sur au moins deux EC2 instances, mais vous pouvez les exécuter HBase sur un seul nœud à des fins d'évaluation.
- Cluster de longue durée : HBase ne s'exécute que sur des clusters de longue durée. Par défaut, la EMR console CLI et Amazon créent des clusters de longue durée.

- Un jeu de paires de EC2 clés Amazon (recommandé) : pour utiliser le protocole réseau Secure Shell (SSH) pour vous connecter au nœud principal et exécuter des commandes HBase shell, vous devez utiliser une paire de EC2 clés Amazon lors de la création du cluster.
- Versions correctes AMI et Hadoop : les HBase clusters ne sont actuellement pris en charge que sur Hadoop 20.205 ou version ultérieure.
- Ganglia (facultatif) : pour surveiller les indicateurs de HBase performance, installez Ganglia lors de la création du cluster.
- Un compartiment Amazon S3 pour les journaux (facultatif) : les journaux pour HBase sont disponibles sur le nœud principal. Si vous voulez que ces journaux soient copiés dans Amazon S3, spécifiez un compartiment S3 qui contiendra les fichiers journaux lorsque vous créez le cluster.

Création d'un cluster avec HBase

Le tableau suivant répertorie les options disponibles lorsque vous utilisez la console pour créer un cluster à HBase l'aide d'une EMR AMI version publiée par Amazon.

Champ	Action
Restore from backup (Rétablir à partir d'une sauvegarde)	Spécifiez s'il faut précharger le HBase cluster avec les données stockées dans Amazon S3.
Backup location (Emplacement de sauvegarde)	Spécifiez l'URI emplacement de la sauvegarde à restaurer dans Amazon S3.
Backup version (Version de sauvegarde)	Éventuellement, spécifiez le nom de version de la sauvegarde dans Backup Location (Emplacement de sauvegarde) à utiliser. Si vous laissez ce champ vide, Amazon EMR utilise la dernière sauvegarde de Backup Location pour remplir le nouveau HBase cluster.
Schedule Regular Backups (Planifier des	Spécifiez s'il convient de planifier des sauvegardes incrémentielles automatiques. La première sauvegarde est une sauvegarde complète pour créer une référence pour les futures sauvegardes incrémentielles.

Champ	Action
sauvegardes régulières)	
Consistent backup (Sauvegarde uniforme)	Spécifiez si les sauvegardes doivent être cohérentes. Une sauvegarde cohérente est une sauvegarde qui interrompt les opérations d'écriture au cours de la phase initiale de sauvegarde, pour la synchronisation entre les nœuds. Les éventuelles opérations d'écriture ainsi interrompues sont placées dans une file d'attente et reprises lorsque la synchronisation est terminée.
Backup frequency (Fréquence de sauvegarde)	Nombre de jours/d'heures/de minutes entre les sauvegardes planifiées.
Backup location (Emplacement de sauvegarde)	L'Amazon S3 URI où les sauvegardes sont stockées. L'emplacement de sauvegarde de chaque HBase cluster doit être différent pour garantir que les sauvegardes différentielles restent correctes.
Backup start time (Heure de début de sauvegarde)	Spécifiez quand la première sauvegarde doit avoir lieu. Vous pouvez définir ce paramètre surnow, ce qui entraîne le démarrage de la première sauvegarde dès que le cluster est en cours d'exécution, ou saisir une date et une heure au ISOformat . Par exemple, 2012-06-15T 20:00 Z fixerait l'heure de début au 15 juin 2012 à 20 heures. UTC

L'exemple de AWS CLI commande suivant lance un cluster avec HBase d'autres applications :

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
  --applications Name=Hue Name=Hive Name=Pig Name=HBase \
  --use-default-roles --ec2-attributes KeyName=myKey \
```

```
--instance-type c1.xlarge --instance-count 3 --termination-protected
```

Une fois la connexion établie entre le Hive et les HBase clusters (comme indiqué dans la procédure précédente), vous pouvez accéder aux données stockées sur le HBase cluster en créant une table externe dans Hive.

L'exemple suivant, lorsqu'il est exécuté à partir de l'invite Hive, crée une table externe qui fait référence aux données stockées dans une HBase table appelée `inputTable`. Vous pouvez ensuite faire référence `inputTable` dans les instructions Hive pour interroger et modifier les données stockées dans le HBase cluster.

Note

L'exemple suivant utilise `protobuf-java-2.4.0a.jar` dans la version AMI 2.3.3, mais vous devez le modifier pour qu'il corresponde à votre version. Pour vérifier la version des Protocol Buffers dont JAR vous disposez, exécutez la commande à l'invite de commande Hive : `! ls /home/hadoop/lib;`

```
add jar lib/emr-metrics-1.0.jar ;
    add jar lib/protobuf-java-2.4.0a.jar ;

set hbase.zookeeper.quorum=ec2-107-21-163-157.compute-1.amazonaws.com ;

create external table inputTable (key string, value string)
    stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
    with serdeproperties ("hbase.columns.mapping" = ":key,f1:col1")
    tblproperties ("hbase.table.name" = "t1");

select count(*) from inputTable ;
```

Personnalisation de la configuration HBase

Bien que les paramètres par défaut devraient fonctionner pour la plupart des applications, vous avez la possibilité de modifier vos paramètres HBase de configuration. Pour ce faire, exécutez l'un des deux scripts d'action d'amorçage :

- `configure-hbase-daemons`—Configure les propriétés des démons `master`, `regionserver` et `zookeeper`. Ces propriétés incluent la taille du segment de mémoire et les options à transmettre à

la machine virtuelle Java (JVM) au démarrage du HBase démon. Vous définissez ces propriétés en tant qu'arguments dans l'action d'amorçage. Cette action de démarrage modifie le fichier de configuration `/home/hadoop/conf/hbase-user-env.sh` sur le cluster. HBase

- `configure-hbase` —Configure les paramètres HBase spécifiques au site, tels que le port auquel le HBase maître doit se lier et le nombre maximal de fois que le client doit réessayer une action. CLI Vous pouvez les one-by-one définir en tant qu'arguments dans l'action bootstrap, ou vous pouvez spécifier l'emplacement d'un fichier de XML configuration dans Amazon S3. Cette action d'amorçage modifie le fichier de configuration `/home/hadoop/conf/hbase-site.xml` sur le HBase cluster.

Note

Ces scripts, comme les autres actions d'amorçage, ne peuvent être exécutés que lorsque le cluster est créé ; vous ne pouvez pas les utiliser pour modifier la configuration d'un HBase cluster en cours d'exécution.

Lorsque vous exécutez les actions `configure-hbase` ou `configure-hbase-daemonsbootstrap`, les valeurs que vous spécifiez remplacent les valeurs par défaut. Toutes les valeurs que vous ne définissez pas explicitement reçoivent les valeurs par défaut.

La configuration HBase à l'aide de ces actions bootstrap est analogue à l'utilisation des actions bootstrap dans Amazon EMR pour configurer les paramètres Hadoop et les propriétés du démon Hadoop. La différence est qu'il HBase n'y a pas d'options de mémoire par processus. Au lieu de cela, les options de mémoire sont définies à l'aide de l'option `--daemon-opts` argument, où *daemon* est remplacé par le nom du démon à configurer.

Configurer les HBase démons

Amazon EMR fournit une action `bootstraps3://region.elasticmapreduce/bootstrap-actions/configure-hbase-daemons`, que vous pouvez utiliser pour modifier la configuration des HBase démons, où *region* est la région dans laquelle vous lancez votre HBase cluster.

Pour configurer HBase des démons à l'aide de AWS CLI, ajoutez l'action `configure-hbase-daemons bootstrap` lorsque vous lancez le cluster afin de configurer un ou plusieurs démons. HBase Vous pouvez définir les propriétés suivantes :

Propriété	Description
<code>hbase-master-opts</code>	Options qui contrôlent la façon dont le démon principal est JVM exécuté. Si cette option est définie, elles remplacent les OPTS variables HBASE _ MASTER _ par défaut.
<code>regionserver-opts</code>	Options qui contrôlent le mode d'JVM exécution du démon du serveur régional. Si cette option est définie, elles remplacent les OPTS variables HBASE _ REGIONSERVER _ par défaut.
<code>zookeeper-opts</code>	Options qui contrôlent la façon dont le JVM démon zookeeper est exécuté. Si cette option est définie, elles remplacent les OPTS variables HBASE _ ZOOKEEPER _ par défaut.

Pour plus d'informations sur ces options, consultez le [fichier `hbase-env.sh`](#) dans la HBase documentation.

Une action d'amorçage permettant de configurer les valeurs de `zookeeper-opts` et `hbase-master-opts` est illustrée dans l'exemple suivant.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hbase-
daemons,\
Args=["--hbase-zookeeper-opts=-Xmx1024m -XX:GCTimeRatio=19", "--hbase-master-opts=-
Xmx2048m", "--hbase-regionserver-opts=-Xmx4096m"]
```

Configuration des paramètres HBase du site

Amazon EMR fournit une action `bootstraps3://elasticmapreduce/bootstrap-actions/configure-hbase`, que vous pouvez utiliser pour modifier la configuration de HBase. Vous pouvez définir des valeurs de configuration one-by-one, comme arguments dans l'action bootstrap, ou vous pouvez spécifier l'emplacement d'un fichier de XML configuration dans Amazon S3. La définition de valeurs de configuration one-by-one est utile si vous n'avez besoin de définir que quelques paramètres de configuration. Il est utile de les définir à l'aide d'un XML fichier si vous devez apporter de nombreuses modifications ou si vous souhaitez enregistrer vos paramètres de configuration pour les réutiliser.

Note

Vous pouvez préfixer le nom du compartiment Amazon S3 avec un préfixe de région, tel que `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase`, où *region* est la région dans laquelle vous lancez votre HBase cluster.

Cette action d'amorçage modifie le fichier `/home/hadoop/conf/hbase-site.xml` de configuration du HBase cluster. L'action bootstrap ne peut être exécutée que lorsque le HBase cluster est lancé.

Pour plus d'informations sur les paramètres du HBase site que vous pouvez configurer, consultez la section [Configuration par défaut](#) dans la HBase documentation.

Définissez l'action `configure-hbase bootstrap` lorsque vous lancez le HBase cluster et spécifiez les valeurs `hbase-site.xml` à modifier.

Pour définir les paramètres individuels HBase du site à l'aide du AWS CLI

- Pour modifier le `hbase.hregion.max.filesize` paramètre, tapez la commande suivante et remplacez *myKey* avec le nom de votre paire de EC2 clés Amazon.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \  
--applications Name=Hue Name=Hive Name=Pig Name=HBase \  
--use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type c1.xlarge --instance-count 3 --termination-protected \  
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-  
hbase,Args=["-s", "hbase.hregion.max.filesize=52428800"]
```

Pour définir les paramètres HBase du site avec un XML fichier à l'aide du AWS CLI

1. Créez une version personnalisée de `hbase-site.xml`. Votre fichier personnalisé doit être valide XML. Pour réduire le risque d'introduction d'erreurs, commencez par la copie par défaut `dehbase-site.xml`, située sur le nœud EMR HBase principal Amazon à l'adresse `/home/hadoop/conf/hbase-site.xml`, et modifiez une copie de ce fichier au lieu de créer un fichier à partir de zéro. Vous pouvez donner un nouveau nom à votre nouveau fichier ou conserver le nom `hbase-site.xml`.
2. Chargez votre fichier `hbase-site.xml` personnalisé sur un compartiment Amazon S3. Les autorisations doivent être définies pour que le AWS compte qui lance le cluster puisse accéder au fichier. Si le AWS compte qui lance le cluster possède également le compartiment Amazon S3, il y a accès.
3. Définissez l'action de démarrage `configure-hbase` lorsque vous lancez le HBase cluster et incluez l'emplacement de votre fichier personnalisé. `hbase-site.xml` L'exemple suivant définit les valeurs de configuration du HBase site selon celles spécifiées dans le fichiers `s3://mybucket/my-hbase-site.xml`. Tapez la commande suivante, remplacez `myKey` avec le nom de votre paire de EC2 clés, et remplacez `mybucket` avec le nom de votre compartiment Amazon S3.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \  
--applications Name=Hue Name=Hive Name=Pig Name=HBase \  
--use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type c1.xlarge --instance-count 3 --termination-protected \  
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-  
hbase,Args=["-s", "hbase.hregion.max.filesize=52428800"]
```

```
--use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type c1.xlarge --instance-count 3 --termination-protected \  
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-  
hbase,Args=["--site-config-file", "s3://mybucket/config.xml"]
```

Si vous spécifiez plusieurs options pour personnaliser le HBase fonctionnement, vous devez ajouter un commutateur d'options à chaque paire clé-valeur, comme indiqué dans l'exemple suivant :

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-  
hbase,Args=["-s", "zookeeper.session.timeout=60000"]
```

Une fois le proxy défini et la SSH connexion ouverte, vous pouvez afficher l'interface utilisateur en ouvrant une fenêtre de navigateur avec `http://master-public-dns-name:60010/master-status`, où *master-public-dns-name* est l'adresse DNS publique du nœud maître du HBase cluster.

Vous pouvez consulter les HBase journaux actuels en utilisant SSH pour vous connecter au nœud principal et en accédant au `mnt/var/log/hbase` répertoire. Ces journaux ne sont pas disponibles une fois que le cluster a été arrêté sauf si vous activez la journalisation dans Amazon S3 lorsque le cluster est lancé.

Sauvegarder et restaurer HBase

Amazon EMR permet de sauvegarder vos HBase données sur Amazon S3, soit manuellement, soit selon un calendrier automatisé. Vous pouvez effectuer des sauvegardes complètes et incrémentielles. Une fois que vous disposez d'une version sauvegardée des HBase données, vous pouvez restaurer cette version dans un HBase cluster. Vous pouvez effectuer une restauration sur un HBase cluster en cours d'exécution ou lancer un nouveau cluster prérempli de données sauvegardées.

Pendant le processus de sauvegarde, HBase continue d'exécuter les commandes d'écriture. Cela garantit la disponibilité du cluster tout au long de la sauvegarde, mais cela entraîne un risque d'incohérence entre les données en cours de sauvegarde et les éventuelles opérations d'écriture en cours d'exécution en parallèle. Pour comprendre les incohérences qui peuvent survenir, vous devez tenir compte du fait que les opérations d'écriture sont réparties entre les nœuds de son cluster. Si une opération d'écriture se produit après qu'un nœud particulier a été interrogé, ces données ne

sont pas incluses dans l'archive de sauvegarde. Vous pouvez même constater que les écritures antérieures sur le HBase cluster (envoyées à un nœud qui a déjà été interrogé) ne figurent peut-être pas dans l'archive de sauvegarde, alors que les écritures ultérieures (envoyées à un nœud avant qu'il ne soit interrogé) sont incluses.

Si une sauvegarde cohérente est requise, vous devez suspendre les écritures HBase pendant la partie initiale du processus de sauvegarde, c'est-à-dire la synchronisation entre les nœuds. Pour cela, vous pouvez spécifier le paramètre `--consistent` lorsque vous demandez une sauvegarde. Avec ce paramètre, les écritures intervenant au cours de cette période sont mises en file d'attente et exécutées dès que la synchronisation se termine. Vous pouvez également planifier des sauvegardes récurrentes, ce qui résout les incohérences au fil du temps, car les données qui manquent dans un cycle de sauvegarde sont sauvegardées lors du cycle suivant.

Lorsque vous sauvegardez HBase des données, vous devez spécifier un répertoire de sauvegarde différent pour chaque cluster. Un moyen simple pour cela consiste à utiliser l'identifiant du cluster dans le chemin d'accès spécifié pour le répertoire de sauvegarde. Par exemple, `s3://mybucket/backups/j-3AEXXXXXX16F2`. Cela garantit que toutes les futures sauvegardes incrémentielles font référence au HBase cluster approprié.

Lorsque vous êtes prêt à supprimer les anciens fichiers de sauvegarde devenus inutiles, nous vous recommandons d'effectuer d'abord une sauvegarde complète de vos HBase données. Ceci garantit la conservation de toutes les données et fournit un point de départ pour les futures sauvegardes incrémentielles. Une fois la sauvegarde complète terminée, vous pouvez accéder à l'emplacement de sauvegarde et supprimer manuellement les anciens fichiers de sauvegarde.

Le processus HBase de sauvegarde utilise S3 DistCp pour l'opération de copie, qui comporte certaines limites concernant l'espace de stockage temporaire des fichiers.

Sauvegarde et restauration à HBase l'aide de la console

La console permet de lancer un nouveau cluster et de le remplir avec les données d'une HBase sauvegarde précédente. Il vous permet également de planifier des sauvegardes incrémentielles périodiques des HBase données. Des fonctionnalités de sauvegarde et de restauration supplémentaires, telles que la possibilité de restaurer des données sur un cluster déjà actif, d'effectuer des sauvegardes manuelles et de planifier des sauvegardes complètes automatisées, sont disponibles à l'aide du CLI.

Pour remplir un nouveau cluster avec des HBase données archivées à l'aide de la console

1. Accédez à la nouvelle EMR console Amazon et sélectionnez Basculer vers l'ancienne console dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).
2. Choisissez Créer un cluster.
3. Dans la section Configuration logicielle, pour Applications supplémentaires, sélectionnez Configurer HBase et ajouter.
4. Dans la boîte de dialogue Add Application (Ajouter une application), cochez Restore From Backup (Restituer à partir d'une sauvegarde).
5. Pour Backup Location, spécifiez l'emplacement de la sauvegarde à charger dans le nouveau HBase cluster. Il doit s'agir d'un Amazon S3 URL du formulaire s3://myawsbucket/backups/.
6. Pour Backup Version (Version de sauvegarde), vous avez la possibilité de spécifier le nom d'une version de sauvegarde à charger en définissant une valeur. Si vous ne définissez pas de valeur pour la version de sauvegarde, Amazon EMR charge la dernière sauvegarde à l'emplacement indiqué.
7. Choisissez Ajouter, puis créez le cluster avec d'autres options, si vous le souhaitez.

Pour planifier des sauvegardes automatisées des HBase données à l'aide de la console

1. Dans la section Configuration logicielle, pour Applications supplémentaires, sélectionnez Configurer HBase et ajouter.
2. Choisissez Schedule Regular Backups (Planifier des sauvegardes régulières).
3. Spécifiez si les sauvegardes doivent être cohérentes. Une sauvegarde cohérente est une sauvegarde qui interrompt les opérations d'écriture au cours de la phase initiale de sauvegarde, pour la synchronisation entre les nœuds. Les éventuelles opérations d'écriture ainsi interrompues sont placées dans une file d'attente et reprises lorsque la synchronisation est terminée.
4. Définissez la fréquence à laquelle les sauvegardes doivent intervenir en entrant un nombre pour Backup Frequency (Fréquence de sauvegarde) et en choisissant Jours, Heures ou Minutes. La première sauvegarde automatique exécutée est une sauvegarde complète ; Amazon EMR enregistre ensuite les sauvegardes incrémentielles selon le calendrier que vous spécifiez.

5. Spécifiez l'emplacement dans Amazon S3 où les sauvegardes doivent être stockées. Chaque HBase cluster doit être sauvegardé dans un emplacement distinct dans Amazon S3 afin de garantir que les sauvegardes incrémentielles sont correctement calculées.
6. Spécifiez quand la première sauvegarde doit intervenir en définissant une valeur pour Backup Start Time (Heure de début de sauvegarde). Vous pouvez définir ce paramètre surnow, ce qui entraîne le démarrage de la première sauvegarde dès que le cluster est en cours d'exécution, ou saisir une date et une heure au [ISOformat](#). Par exemple, 2013-09-26T 20:00 Z définit l'heure de début au 26 septembre 2013 à 20 heures. UTC
7. Choisissez Ajouter.
8. Procédez à la création du cluster avec d'autres options, si vous le souhaitez.

Moniteur HBase avec CloudWatch

Amazon EMR indique trois indicateurs CloudWatch que vous pouvez utiliser pour surveiller vos HBase sauvegardes. Ces mesures sont poussées CloudWatch à des intervalles de cinq minutes et sont fournies gratuitement.

Métrique	Description
HBaseBackupFailed	<p>Si la dernière sauvegarde a échoué. La valeur est définie sur 0 par défaut et mise à jour sur 1 en cas d'échec de la tentative de sauvegarde précédente. Cette métrique n'est signalée que pour les HBase clusters.</p> <p>Cas d'utilisation : surveiller les HBase sauvegardes</p> <p>Unités : nombre</p>
HBaseMostRecentBackupDuration	<p>Délai nécessaire à l'exécution de la précédente sauvegarde. Cette métrique est définie même si la dernière sauvegarde a réussi ou a échoué. Lorsque la sauvegarde est en cours, cette métrique retourne le nombre de minutes qui se sont écoulées depuis le démarrage de la sauvegarde. Cette métrique n'est signalée que pour les HBase clusters.</p> <p>Cas d'utilisation : Surveiller les HBase sauvegardes</p>

Métrique	Description
	Unités : minutes
HBaseTimeSinceLastSuccessfulBackup	<p>Nombre de minutes écoulées après le démarrage de la dernière HBase sauvegarde réussie sur votre cluster. Cette métrique n'est signalée que pour les HBase clusters.</p> <p>Cas d'utilisation : surveiller les HBase sauvegardes</p> <p>Unités : minutes</p>

Configurer Ganglia pour HBase

Vous configurez Ganglia pour HBase utiliser l'action `configure-hbase-for-gangliabootstrap`. Cette action de bootstrap est configurée HBase pour publier des métriques sur Ganglia.

Vous devez configurer HBase et Ganglia lorsque vous lancez le cluster ; les rapports Ganglia ne peuvent pas être ajoutés à un cluster en cours d'exécution.

Ganglia stocke également les fichiers journaux sur le serveur à l'adresse `/mnt/var/log/ganglia/irids`. Si vous avez configuré votre cluster pour conserver les fichiers journaux dans un compartiment Amazon S3, les fichiers journaux Ganglia y sont conservés également.

Pour lancer un cluster avec Ganglia forHBase, utilisez l'action `configure-hbase-for-gangliabootstrap` comme indiqué dans l'exemple suivant.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Test cluster" --ami-version 3.3 \
--applications Name=Hue Name=Hive Name=Pig Name=HBase Name=Ganglia \
--use-default-roles --ec2-attributes KeyName=myKey \
--instance-type c1.xlarge --instance-count 3 --termination-protected \
```

```
--bootstrap-actions Path=s3://elasticmapreduce/bootstrap-actions/configure-hbase-for-ganglia
```

Une fois que le cluster a été lancé et Ganglia configuré, vous pouvez accéder aux graphiques et rapports Ganglia à l'aide de l'interface graphique en cours d'exécution sur le nœud maître.

Spécificités de l'application Pig pour AMI les versions antérieures d'Amazon EMR

Versions de Pig prises en charge

La version de Pig que vous pouvez ajouter à votre cluster dépend de la version d'Amazon EMR AMI et de la version de Hadoop que vous utilisez. Le tableau ci-dessous indique quelles AMI versions et versions de Hadoop sont compatibles avec les différentes versions de Pig. Nous recommandons d'utiliser la dernière version de Pig disponible pour tirer parti des améliorations de performances et des nouvelles fonctionnalités.

Lorsque vous utilisez le API pour installer Pig, la version par défaut est utilisée sauf si vous spécifiez `--pig-versions` comme argument l'étape qui charge Pig sur le cluster lors de l'appel à [RunJobFlow](#).

Version de Pig	AMIversion	Paramètres de configuration	Détails de la version de Pig
0.12.0 Notes de mise à jour Documentation	3.1.0 et ultérieures	<pre>--ami-version 3.1</pre> <pre>--ami-version 3.2</pre> <pre>--ami-version 3.3</pre>	<p>Ajoute la prise en charge des éléments suivants :</p> <ul style="list-style-type: none"> • Streaming UDFs sans JVM implémentations • ASSERTet opérateurs IN • CASEexpression • AvroStorage en tant que fonction intégrée à Pig.

Version de Pig	AMIversion	Paramètres de configuration	Détails de la version de Pig
			<ul style="list-style-type: none"> ParquetLoader et ParquetStorer en tant que fonctions intégrées BigInteger et BigDecimal types
0.11.1.1 Notes de mise à jour Documentation	2.2 et ultérieures	<pre>--pig-versions 0.11.1.1 --ami-version 2.2</pre>	Améliore les performances de la LOAD commande PigStorage si l'entrée réside dans Amazon S3.
0.11.1 Notes de mise à jour Documentation	2.2 et ultérieures	<pre>--pig-versions 0.11.1 --ami-version 2.2</pre>	Ajoute la prise en charge de JDK 7, Hadoop 2, des fonctions définies par l'utilisateur de Groovy, de l' SchemaTup le optimisation, de nouveaux opérateurs, etc. Pour plus d'informations, consultez Journal de modifications Pig 0.11.1 .
0.9.2.2 Notes de mise à jour Documentation	2.2 et ultérieures	<pre>--pig-versions 0.9.2.2 --ami-version 2.2</pre>	Ajoute une prise en charge pour Hadoop 1.0.3.

Version de Pig	AMIversion	Paramètres de configuration	Détails de la version de Pig
0.9.2.1 Notes de mise à jour Documentation	2.2 et ultérieures	--pig-versions 0.9.2.1 --ami-version 2.2	Ajoute la prise en charge de MapR.
0.9.2 Notes de mise à jour Documentation	2.2 et ultérieures	--pig-versions 0.9.2 --ami-version 2.2	Inclut plusieurs améliorations de performances et correctifs. Pour obtenir des informations détaillées sur les modifications apportées dans la version Pig 0.9.2, consultez Journal de modifications Pig 0.9.2 .
0.9.1 Notes de mise à jour Documentation	2.0	--pig-versions 0.9.1 --ami-version 2.0	
0.6 Notes de mise à jour	1.0	--pig-versions 0.6 --ami-version 1.0	
0.3 Notes de mise à jour	1.0	--pig-versions 0.3 --ami-version 1.0	

Détails de la version de Pig

Amazon EMR prend en charge certaines versions de Pig auxquelles des EMR correctifs Amazon supplémentaires peuvent être appliqués. Vous pouvez configurer la version de Pig à exécuter sur les EMR clusters Amazon. Pour plus d'informations sur cette étape, consultez [Apache Pig](#). Les sections suivantes décrivent les différentes versions de Pig et les correctifs appliqués aux versions chargées sur AmazonEMR.

Correctifs Pig

Cette section décrit les correctifs personnalisés appliqués aux versions de Pig disponibles sur AmazonEMR.

Correctifs Pig 0.11.1.1

La EMR version Amazon de Pig 0.11.1.1 est une version de maintenance qui améliore les performances des LOAD commandes PigStorage si l'entrée réside dans Amazon S3.

Correctifs Pig 0.11.1

La EMR version Amazon de Pig 0.11.1 contient toutes les mises à jour fournies par l'Apache Software Foundation et les EMR correctifs Amazon cumulés de la version 0.9.2.2 de Pig. Cependant, il n'y a aucun nouveau correctif EMR spécifique à Amazon dans Pig 0.11.1.

Correctifs Pig 0.9.2

Apache Pig 0.9.2 est une version de maintenance de Pig. L'EMRéquipe Amazon a appliqué les correctifs suivants à la EMR version Amazon de Pig 0.9.2.

Correctif	Description
PIG-1429	<p>Ajout du type de données Boolean à Pig en tant que type de données de première classe. Pour plus d'informations, rendez-vous sur https://issues.apache.org/jira/browse/ PIG -1429.</p> <p>Statut : validé</p> <p>Correction dans la version Apache Pig : 0.10</p>
PIG-1824	<p>Support des modules d'importation dans JythonUDF. Pour plus d'informations, rendez-vous sur https://issues.apache.org/jira/browse/ PIG -1824.</p>

Correctif	Description
	Statut : validé Correction dans la version Apache Pig : 0.10
PIG-2010	JARsBundle enregistré dans le cache distribué. Pour plus d'informations, rendez-vous sur https://issues.apache.org/jira/browse/PIG-2010 . Statut : validé Correction dans la version Apache Pig : 0.11
PIG-2456	Ajout d'un fichier ~/.pigbootup dans lequel l'utilisateur peut spécifier des déclarations Pig par défaut. Pour plus d'informations, rendez-vous sur https://issues.apache.org/jira/browse/PIG-2456 . Statut : validé Correction dans la version Apache Pig : 0.11
PIG-2623	Support à l'utilisation des chemins Amazon S3 pour l'enregistrementUDFs. Pour plus d'informations, rendez-vous sur https://issues.apache.org/jira/browse/PIG-2623 . Statut : validé Correction dans les versions Apache Pig : 0.10, 0.11

Correctifs Pig 0.9.1

L'EMRéquipe Amazon a appliqué les correctifs suivants à la EMR version Amazon de Pig 0.9.1.

Correctif	Description
Support JAR des fichiers et des scripts Pig dans DFS	Ajoutez la prise en charge de l'exécution de scripts et de l'enregistrement de JAR fichiers stockés dans HDFS Amazon S3 ou dans d'autres systèmes de fichiers distribués. Pour plus

Correctif	Description
	<p>d'informations, rendez-vous sur https://issues.apache.org/jira/browse/ PIG -1505.</p> <p>Statut : validé</p> <p>Correction dans la version Apache Pig : 0.8.0</p>
<p>Prise en charge de plusieurs systèmes de fichiers dans Pig</p>	<p>Ajout de la prise en charge des scripts Pig pour lire les données d'un système de fichiers et les écrire dans un autre. Pour plus d'informations, rendez-vous sur https://issues.apache.org/jira/browse/ PIG -1564.</p> <p>Statut : non validé</p> <p>Correction dans la version Apache Pig : non applicable</p>
<p>Ajouter la date, l'heure et la chaîne de caractères de Piggybank. UDFs</p>	<p>Ajoutez la date, l'heure et la chaîne UDFs pour prendre en charge les scripts Pig personnalisés. Pour plus d'informations, rendez-vous sur https://issues.apache.org/jira/browse/ PIG -1565.</p> <p>Statut : non validé</p> <p>Correction dans la version Apache Pig : non applicable</p>

Clusters Pig en mode interactif ou en mode de traitement par lots

Amazon vous EMR permet d'exécuter des scripts Pig selon deux modes :

- Interactive
- Par lots

Lorsque vous lancez un cluster de longue durée à l'aide de la console ou du AWS CLI, vous pouvez vous connecter au nœud principal ssh en tant qu'utilisateur Hadoop et utiliser le shell Grunt pour développer et exécuter vos scripts Pig de manière interactive. Utiliser Pig de façon interactive vous permet de réviser plus facilement le script Pig qu'en mode de traitement par lots. Lorsque vous avez révisé le script Pig en mode interactif et que tout est correct, vous pouvez charger le script sur

Amazon S3 et utiliser le mode de traitement par lots pour exécuter le script en production. Vous pouvez également soumettre des commandes Pig de façon interactive sur un cluster en cours d'exécution pour analyser et transformer les données en fonction des besoins.

En mode de traitement par lots, vous chargez votre script Pig sur Amazon S3, puis soumettez le travail au cluster dans le cadre d'une étape. Les étapes Pig peuvent être soumises à un cluster de longue durée ou à un cluster transitoire.

Spécificités de l'application Spark avec AMI les versions antérieures d'Amazon EMR

Utilisation interactive de Spark ou en mode de traitement par lots

Amazon vous EMR permet d'exécuter des applications Spark selon deux modes :

- Interactive
- Par lots

Lorsque vous lancez un cluster de longue durée à l'aide de la console ou du AWS CLI, vous pouvez vous connecter au nœud principal SSH en tant qu'utilisateur Hadoop et utiliser le shell Spark pour développer et exécuter vos applications Spark de manière interactive. L'utilisation de Spark vous permet de façon interactive de créer un prototype ou de tester les applications Spark plus facilement que dans un environnement de traitement par lots. Une fois que vous avez correctement révisé l'application Spark en mode interactif, vous pouvez placer cette application JAR ou ce programme Python dans le système de fichiers local du nœud principal du cluster sur Amazon S3. Vous pouvez alors soumettre l'application en tant que flux de travail de traitement par lots.

En mode de traitement par lots, chargez votre script Spark sur Amazon S3 ou le système de fichiers local du nœud principal, puis soumettez le travail au cluster en tant qu'étape. Les étapes Spark peuvent être soumises à un cluster de longue durée ou à un cluster transitoire.

Création d'un cluster avec Spark installé

Pour lancer un cluster avec Spark installé à l'aide de la console

1. Accédez à la nouvelle EMR console Amazon et sélectionnez Basculer vers l'ancienne console dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).

2. Choisissez Créer un cluster.
3. Pour la configuration logicielle, choisissez la version AMI de publication dont vous avez besoin.
4. Pour Applications to be installed (Applications à installer), choisissez Spark dans la liste, puis choisissez Configurer et ajouter.
5. Ajoutez des arguments pour modifier la configuration de Spark comme vous le souhaitez. Pour de plus amples informations, veuillez consulter [Configuration de Spark](#). Choisissez Ajouter.
6. Sélectionnez d'autres options si nécessaire, puis choisissez Create cluster (Créer le cluster).

L'exemple suivant montre comment créer un cluster avec Spark au moyen de Java :

```
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
SupportedProductConfig sparkConfig = new SupportedProductConfig()
    .withName("Spark");

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("Spark Cluster")
    .withAmiVersion("3.11.0")
    .withNewSupportedProducts(sparkConfig)
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myKeyName")
        .withInstanceCount(1)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m3.xlarge")
        .withSlaveInstanceType("m3.xlarge")
    );
RunJobFlowResult result = emr.runJobFlow(request);
```

Configuration de Spark

Vous configurez Spark lorsque vous créez un cluster en exécutant l'action bootstrap située dans le référentiel [aws-labs/emr-bootstrap-actions /spark](https://github.com/aws-labs/emr-bootstrap-actions/tree/master/spark) sur Github. Pour les arguments acceptés par l'action bootstrap, consultez le [README](#) dans ce référentiel. L'action d'amorçage configure les propriétés dans le fichier `$SPARK_CONF_DIR/spark-defaults.conf`. Pour plus d'informations sur les paramètres, consultez la rubrique Spark Configuration dans la documentation Spark. Vous pouvez remplacer « dernière » dans ce qui suit URL par le numéro de version de Spark que vous installez, par exemple `2.2.0`. <http://spark.apache.org/docs/latest/configuration.html>.

Vous pouvez également configurer Spark dynamiquement au moment de l'envoi de chaque application. Un paramètre permettant d'optimiser automatiquement l'allocation de ressources pour un

exécuteur est disponible dans le fichier de configuration spark. Pour de plus amples informations, veuillez consulter [Remplacement des paramètres de configuration par défaut de Spark](#).

Modification des paramètres Spark par défaut

L'exemple suivant montre comment créer un cluster avec `spark.executor.memory` défini sur 2G au moyen de l' AWS CLI.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Spark cluster" --ami-version 3.11.0 \  
--applications Name=Spark, Args=[-d,spark.executor.memory=2G] --ec2-attributes \  
KeyName=myKey \  
--instance-type m3.xlarge --instance-count 3 --use-default-roles
```

Soumission d'un travail à Spark

Pour soumettre un travail à un cluster, utilisez une étape pour exécuter le `spark-submit` script sur votre EMR cluster. Ajoutez l'étape à l'aide de la `addJobFlowSteps` méthode ci-dessous [AmazonElasticMapReduceClient](#):

```
AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);  
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);  
StepFactory stepFactory = new StepFactory();  
AddJobFlowStepsRequest req = new AddJobFlowStepsRequest();  
req.withJobFlowId("j-1K48XXXXXXHCB");  
  
List<StepConfig> stepConfigs = new ArrayList<StepConfig>();  
  
StepConfig sparkStep = new StepConfig()  
    .withName("Spark Step")  
    .withActionOnFailure("CONTINUE")  
    .withHadoopJarStep(stepFactory.newScriptRunnerStep("/home/hadoop/spark/bin/spark-  
submit", "--class", "org.apache.spark.examples.SparkPi", "/home/hadoop/spark/lib/spark-  
examples-1.3.1-hadoop2.4.0.jar", "10"));
```

```
stepConfigs.add(sparkStep);
req.withSteps(stepConfigs);
AddJobFlowStepsResult result = emr.addJobFlowSteps(req);
```

Remplacement des paramètres de configuration par défaut de Spark

Vous souhaitez probablement remplacer les valeurs de configuration par défaut de Spark indépendamment pour chaque application. Vous pouvez faire cela lorsque vous soumettez des applications à l'aide d'une étape, ce qui transmet essentiellement des options à `spark-submit`. Par exemple, vous pouvez changer la mémoire allouée à un processus d'exécuteur en modifiant `spark.executor.memory`. Vous pouvez fournir le commutateur `--executor-memory` avec un argument tel que :

```
/home/hadoop/spark/bin/spark-submit --executor-memory 1g --class
org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/spark-examples*.jar 10
```

De même, vous pouvez ajuster `--executor-cores` et `--driver-memory`. Dans une étape, vous pouvez fournir les arguments suivants à l'étape :

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/
spark-examples*.jar 10
```

Vous pouvez également ajuster les paramètres qui n'ont peut-être pas de commutateur intégré à l'aide de l'option `--conf`. Pour plus d'informations sur les autres paramètres ajustables, consultez la rubrique [Chargement dynamique de propriétés Spark](#) dans la documentation Apache Spark.

Différences entre l'AMI utilitaire S3 et les versions antérieures d'Amazon EMR

DistCp Versions S3 prises en charge sur Amazon EMR

Les DistCp versions S3 suivantes sont prises en charge dans les EMR AMI versions d'Amazon. Les versions S3 postérieures à 1.0.7 se trouvent directement sur les clusters. Utilisez l'JAR d'entrée `/home/hadoop/lib` pour accéder aux dernières fonctionnalités.

Version	Description	Date de publication
1.0.8	Ajoute les options <code>--appendToLastFile</code> , <code>--requirePreviousManifest</code> et <code>--storageClass</code> .	3 janvier 2014
1.0.7	Ajoute l'option <code>--s3ServerSideEncryption</code> .	2 mai 2013
1.0.6	Ajoute l'option <code>--s3Endpoint</code> .	6 août 2012
1.0.5	Améliore la capacité à spécifier la version de S3 DistCp à exécuter.	27 juin 2012
1.0.4	Améliore l'option <code>--deleteOnSuccess</code> .	19 juin 2012
1.0.3	Ajoute une prise en charge pour les options <code>--numberOfFiles</code> et <code>--startingIndex</code> .	12 juin 2012
1.0.2	Améliore l'affectation des noms lors de l'utilisation de groupes.	6 juin 2012
1.0.1	Version initiale de S3DistCp.	19 janvier 2012

Ajouter une étape de DistCp copie S3 à un cluster

Pour ajouter une étape de DistCp copie S3 à un cluster en cours d'exécution, tapez la commande suivante, remplacez `j-3GYXXXXXX9I0K` avec votre identifiant de cluster, et remplacez `mybucket` avec le nom de votre compartiment Amazon S3.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr add-steps --cluster-id j-3GYXXXXXX9I0K \
```

```
--steps Type=CUSTOM_JAR,Name="S3DistCp step",Jar=/home/hadoop/lib/emr-s3distcp-1.0.jar,
\
Args=["--s3Endpoint,s3-eu-west-1.amazonaws.com",\
"--src,s3://mybucket/logs/j-3GYXXXXXX9I0J/node/",\
"--dest,hdfs:///output",\
"--srcPattern,.*[a-zA-Z,]+" ]
```

Exemple Charger CloudFront les connexions Amazon dans HDFS

Cet exemple charge les connexions CloudFront Amazon en HDFS ajoutant une étape à un cluster en cours d'exécution. Au cours du processus, le format de compression passe de Gzip (CloudFront par défaut) à LZO. Cela est utile car les données compressées à l'aide de Gzip LZO peuvent être divisées en plusieurs cartes lors de leur décompression. Vous n'avez donc pas à attendre que la compression soit terminée, comme vous le faites avec Gzip. Cela permet d'obtenir de meilleures performances lorsque vous analysez les données à l'aide d'Amazon EMR. Cet exemple améliore également les performances en utilisant l'expression régulière spécifiée dans l'option `--groupBy` pour combiner tous les journaux d'une heure donnée dans un seul fichier. Les EMR clusters Amazon sont plus efficaces lorsqu'ils traitent quelques fichiers volumineux LZO compressés que lorsqu'ils traitent de nombreux petits fichiers compressés au format GZIP. Pour diviser LZO des fichiers, vous devez les indexer et utiliser la bibliothèque tierce `hadoop-lzo`.

Pour charger Amazon CloudFront Logs dans HDFS, tapez la commande suivante, remplacez `j-3GYXXXXXX9I0K` avec votre identifiant de cluster, et remplacez `mybucket` avec le nom de votre compartiment Amazon S3.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr add-steps --cluster-id j-3GYXXXXXX9I0K \
--steps Type=CUSTOM_JAR,Name="S3DistCp step",Jar=/home/hadoop/lib/emr-s3distcp-1.0.jar,
\
Args=["--src,s3://mybucket/cf","--dest,hdfs:///local",\
"--groupBy,. *XABCD12345678. ([0-9]+-[0-9]+-[0-9]+-[0-9]+). *",\
"--targetSize,128",
"--outputCodec,lzo","--deleteOnSuccess"]
```

Imaginons le cas où l'exemple précédent est exécuté sur les fichiers CloudFront journaux suivants.

```
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-01.HLUS3JKx.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-01.I9CNAZrg.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.YRRwERSA.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.dshVLXFE.gz
s3://amzn-s3-demo-bucket1/cf/XABCD12345678.2012-02-23-02.LpLfuShd.gz
```

S3 DistCp copie, concatène et compresse les fichiers dans les deux fichiers suivants, le nom du fichier étant déterminé par la correspondance établie par l'expression régulière.

```
hdfs:///local/2012-02-23-01.lzo
hdfs:///local/2012-02-23-02.lzo
```

Nouveautés

Cette page décrit les modifications et les fonctionnalités disponibles dans les dernières versions d'Amazon EMR 7.x, 6.x et 5.x.

Ces notes de mise à jour sont également disponibles sur les pages [Amazon EMR 7.2.0](#), [Amazon EMR 6.15.0](#) et [Amazon EMR 5.36.2](#), ainsi que les versions de l'application, les versions des composants et les classifications de configuration disponibles pour chaque version.

- Pour les notes de publication des versions précédentes, consultez [EMR Archive des notes de publication sur Amazon](#).
- Pour recevoir des mises à jour lorsqu'une nouvelle EMR version d'Amazon est disponible, abonnez-vous au [RSS flux des notes de EMR publication d'Amazon](#).

Note

Les versions ultérieures d'Amazon EMR utilisent AWS Signature Version 4 (SigV4) pour authentifier les demandes adressées à Amazon S3. Nous vous recommandons d'utiliser une EMR version d'Amazon compatible SigV4 afin d'accéder aux nouveaux compartiments S3 et d'éviter toute interruption de vos charges de travail. Pour plus d'informations et pour obtenir la liste des EMR versions d'Amazon compatibles SigV4, consultez [Amazon EMR et AWS Signature Version 4](#).

Amazon EMR 7.2.0 (dernière version de la série 7.x)

Les nouvelles EMR versions d'Amazon sont mises à disposition dans différentes régions sur une période de plusieurs jours, en commençant par la première région à la date de sortie initiale. Il est possible que la dernière version ne soit pas disponible dans votre région pendant cette période.

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 7.2.0 d'Amazon. Les modifications concernent la version 7.2.0.

Nouvelles fonctionnalités

- Mises à niveau des applications – Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0.

- Amazon EMR ajoute un support afin que vous puissiez utiliser d'autres applications telles que HBase Flink et Hive avec la classe de stockage Amazon S3 Express One Zone.
- Cette version ajoute la capacité de lire les objets restaurés, afin que vous puissiez lire les objets Glacier depuis un emplacement S3 avec le S3A protocole. Cette fonctionnalité fonctionne avec Spark, Flink et Hive.
- Étiquettes de nœuds avec dimensionnement géré : utilisez des étiquettes de nœuds avec vos instances en fonction du type de marché ou du type de nœud pour améliorer le dimensionnement automatique d'AmazonEMR. Pour plus d'informations, consultez la section [Utilisation du dimensionnement géré sur Amazon EMR](#).

Problèmes connus

- Python 3.11 n'est pas pris en charge avec EMR Studio.

Modifications, améliorations et problèmes résolus

- Cette version corrige un problème de blocage qui peut survenir lors des opérations internes de nettoyage des étapes. Cette opération gère le cycle de vie des étapes au fur et à mesure qu'elles se terminent sur le EMR cluster. Ce problème affecte les EMR opérations critiques d'Amazon, telles que le fonctionnement des étapes et le dimensionnement.
- Cette version résout un problème AMIs selon lequel les clusters personnalisés dotés de certains fichiers journaux préexistants peuvent provoquer l'échec du démon de gestion des EMR journaux Amazon.
- Amazon EMR 7.2.0 met à niveau le EMR daemon Amazon responsable de la gestion des clusters et des activités de surveillance de la version AWS SDK 1 à la version 2.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2023.52040708.0	6,196-102,177 mai 2023	8 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), Canada Ouest (Calgary), AWS

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			GovCloud (États-Unis Ouest), AWS GovCloud (États-Unis Est), Chine (Pékin), Chine (Ningxia)

Amazon EMR 6.15.0 (dernière version de la série 6.x)

Les nouvelles EMR versions d'Amazon sont mises à disposition dans différentes régions sur une période de plusieurs jours, en commençant par la première région à la date de sortie initiale. Il est possible que la dernière version ne soit pas disponible dans votre région pendant cette période.

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.15.0 d'Amazon. Les modifications ont été apportées à la version 6.14.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.15.0](#).

Nouvelles fonctionnalités

- Mises à niveau des applications –Amazon EMR 6.15.0 application upgrades include Apache Hadoop 3.3.6, Apache Hudi 0.14.0-amzn-0, Iceberg 1.4.0-amzn-0, and Trino 426..
- Des [lancements plus rapides pour les EMR clusters qui s'exécutent sur EC2](#) : le lancement d'un EC2 cluster Amazon EMR on est désormais jusqu'à 35 % plus rapide. Grâce à cette amélioration, la plupart des clients peuvent lancer leurs clusters en moins de 5 minutes.
- [CodeWhisperer pour EMR Studio](#) — Vous pouvez désormais utiliser Amazon CodeWhisperer avec Amazon EMR Studio pour obtenir des recommandations en temps réel lorsque vous écrivez du code JupyterLab. CodeWhisperer peut compléter vos commentaires, terminer des lignes de code individuelles, faire line-by-line des recommandations et générer des fonctions entièrement formées.
- Temps de [redémarrage des tâches plus rapides avec Flink](#) — Avec Amazon EMR 6.15.0 et versions ultérieures, plusieurs nouveaux mécanismes sont disponibles pour Apache Flink afin d'améliorer le temps de redémarrage des tâches lors des opérations de restauration ou de dimensionnement des tâches. Cela permet d'optimiser la vitesse de récupération et de redémarrage des graphes d'exécution afin d'améliorer la stabilité des tâches.

- [Contrôle d'accès détaillé et au niveau des tables pour les formats de tables ouvertes : avec Amazon EMR 6.15.0 et versions ultérieures, lorsque vous exécutez des tâches Spark sur Amazon EMR sur des clusters EC2 qui accèdent aux données du catalogue de données AWS Glue, vous pouvez appliquer des autorisations au niveau des tables, des lignes, des colonnes et des cellules AWS Lake Formation aux tables basées](#) sur Hudi, Iceberg ou Delta Lake.
- **Mise à niveau vers Hadoop** — Amazon EMR 6.15.0 inclut une mise à niveau d'Apache Hadoop vers la version 3.3.6. Hadoop 3.3.6 était la dernière version au moment du déploiement d'Amazon EMR 6.15, publiée par Apache en juin 2023. Les versions précédentes d'Amazon EMR (6.9.0 à 6.14.x) utilisaient Hadoop 3.3.3.

La mise à niveau inclut des centaines d'améliorations et de correctifs, ainsi que des fonctionnalités telles que des paramètres de nœud de données reconfigurables, la DFSAdmin possibilité de lancer des opérations de reconfiguration en masse sur tous les nœuds de données actifs et un vectorisé API qui permet aux lecteurs gourmands en recherche de spécifier plusieurs plages à lire. Hadoop 3.3.6 ajoute également le support HDFS APIs et la sémantique de son log write-ahead (WAL), afin qu'HBase puisse fonctionner sur d'autres implémentations de systèmes de stockage. Pour plus d'informations, consultez les journaux des modifications pour les versions [3.3.4](#), [3.3.5](#) et [3.3.6](#) dans la documentation Apache Hadoop.

- **Support AWS SDK pour Java, version 2** - Les applications Amazon EMR 6.15.0 peuvent utiliser AWS SDK les versions Java [1.12.569](#) ou [2.20.160](#) si l'application prend en charge la version v2. La version 2.x AWS SDK pour Java est une réécriture majeure de la base de code de la version 1.x. Il repose sur Java 8+ et ajoute plusieurs fonctionnalités fréquemment demandées. Il s'agit notamment de la prise en charge des E/S non bloquantes et de la possibilité de connecter une HTTP implémentation différente lors de l'exécution. Pour plus d'informations, y compris un guide SDK de migration de Java v1 à v2, consultez le guide [AWS SDK pour Java, version 2](#).

Modifications, améliorations et problèmes résolus

- Pour améliorer vos EMR clusters à haute disponibilité, cette version permet la connectivité aux EMR daemons Amazon sur des hôtes locaux qui utilisent des points de terminaison. IPv6
- Cette version permet à la version TLS 1.2 de communiquer avec les nœuds ZooKeeper provisionnés sur tous les nœuds principaux de votre cluster de haute disponibilité.
- Cette version améliore la gestion des fichiers journaux de ZooKeeper transactions conservés sur les nœuds principaux afin de minimiser les scénarios dans lesquels les fichiers journaux dépassent les limites et interrompent les opérations du cluster.

- Cette version rend la communication intra-nœud plus résiliente pour les clusters à haute disponibilité EMR. Cette amélioration réduit le risque d'échec des actions d'amorçage ou de démarrage du cluster.
- Dans Amazon EMR 6.15.0, Tez introduit des configurations que vous pouvez spécifier pour ouvrir de manière asynchrone les divisions d'entrée dans une division groupée Tez. Cela permet d'accélérer les performances des requêtes de lecture lorsqu'il existe un grand nombre de divisions d'entrée dans une seule division groupée Tez. Pour plus d'informations, consultez la rubrique [Ouverture fractionnée asynchrone Tez](#).
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap),

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	13 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

Amazon EMR 5.36.2 (dernière version de la série 5.x)

Les nouvelles EMR versions d'Amazon sont mises à disposition dans différentes régions sur une période de plusieurs jours, en commençant par la première région à la date de sortie initiale. Il est possible que la dernière version ne soit pas disponible dans votre région pendant cette période.

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.36.2 d'Amazon. Les modifications sont relatives à la version 5.36.1. Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Modifications, améliorations et problèmes résolus

- Cette version améliore la logique de réduction du cluster afin qu'Amazon EMR ne réduise pas les nœuds principaux en dessous du paramètre de facteur de HDFS réplication défini pour le cluster. Cette amélioration répond aux exigences de redondance des données et réduit le risque de blocage d'une opération de dimensionnement.
- Cette version ajoute un nouveau mécanisme de nouvelle tentative au flux de travail de dimensionnement des clusters pour ceux qui exécutent Presto ou Trino. Cette amélioration réduit le risque que le redimensionnement du cluster s'exécute indéfiniment en raison de l'échec d'une seule opération de redimensionnement. Cela améliore également l'utilisation du cluster, car celui-ci augmente et diminue la capacité plus rapidement.
- Résout un problème selon lequel les opérations de réduction du cluster pouvaient être bloquées pendant qu'Amazon mettait hors service EMR gracieusement un nœud principal et que celui-ci devenait inutilisable avant sa mise hors service complète.
- Améliore la stabilité d'un nœud dans un cluster à haute disponibilité comportant plusieurs nœuds principaux lorsqu'Amazon EMR redémarre un seul nœud.

- Optimise la gestion des journaux avec Amazon EMR exécuté sur AmazonEC2. C'est pourquoi vous constaterez peut-être une légère réduction des coûts de stockage pour les journaux de votre cluster.
- Améliore la gestion des fichiers journaux de ZooKeeper transactions conservés sur les nœuds principaux afin de minimiser les scénarios dans lesquels les fichiers journaux dépassent les limites et interrompent les opérations du cluster.
- Corrige un bogue rare qui peut provoquer l'échec d'un cluster à haute disponibilité comportant plusieurs nœuds principaux en raison de l'impossibilité de communiquer avec le Yarn ResourceManager.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			<p>, Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)</p>

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 503.0	4,14,343	xxxxxx, 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

Amazon EMR et AWS Signature Version 4

EMR Les versions Amazon utilisent AWS Signature Version 4 (SigV4) pour authentifier les demandes adressées à Amazon S3. Les compartiments créés dans Amazon S3 après le 24 juin 2020 ne

prennent pas en charge les demandes signées par Signature Version 2 (SigV2). Les buckets créés le 24 juin 2020 ou avant cette date continueront de prendre en charge le protocole SIGv2. Nous vous recommandons de migrer vers une EMR version d'Amazon compatible SigV4 afin de pouvoir accéder à de nouveaux compartiments S3 et d'éviter toute interruption de vos charges de travail.

Si vous utilisez des applications incluses dans Amazon, EMR telles qu'Apache Spark, Apache Hive et Presto, vous n'avez pas besoin de modifier le code de votre application pour l'utiliser. SigV4 Si vous utilisez des applications personnalisées qui ne sont pas incluses dans AmazonEMR, vous devrez peut-être mettre à jour votre code pour pouvoir les utiliser SigV4. Pour plus d'informations, consultez la section [Passage de Signature Version 2 à Signature Version 4](#) dans le Guide de l'utilisateur Amazon S3.

Les EMR versions Amazon suivantes sont prises en charge SigV4 :emr-4.7.4, emr-4.8.5, emr-4.9.6, emr-4.10.1, emr-5.1.1, emr-5.2.3, emr-5.3.2, emr-5.4.1, emr-5.5.4, emr-5.6.1, emr-5.7.1, emr-5.8.3, emr-5.9.1, emr-5.10.1, emr-5.11.4, emr-5.12.3, emr-5.13.1, emr-5.14.2, emr-5.15.1, emr-5.16.1, emr-5.17.2, emr-5.18.1, emr-5.19.1, emr-5.20.1, emr-5.21.2, and emr-5.22.0 and higher. Toutes les versions 6.x et 7.x prennent en charge SigV4.

Approche d'atténuation CVE -2021-44228

Note

Pour les EMR versions 6.9.0 et ultérieures d'Amazon, tous les composants installés par Amazon EMR qui utilisent les bibliothèques Log4j utilisent Log4j version 2.17.1 ou ultérieure.

Amazon EMR s'exécute sur EC2

Le problème abordé dans [CVE-2021-44228](#) concerne les versions principales d'Apache Log4j comprises entre 2.0.0 et 2.14.1 lors du traitement d'entrées provenant de sources non fiables. EMR Les clusters Amazon lancés avec les versions EMR 5.x jusqu'à 5.34.0 et EMR 6.x jusqu'à Amazon EMR 6.5.0 incluent des frameworks open source tels qu'Apache Hive, Flink, Presto et Trino, qui utilisent ces versions d'Apache Log4j. HUDI Cependant, de nombreux clients utilisent les frameworks open source installés sur leurs EMR clusters Amazon pour traiter et enregistrer les entrées provenant de sources non fiables.

Nous vous recommandons d'appliquer la « Amazon EMR Bootstrap Action Solution for Log4j CVE -2021-44228 » comme décrit dans la section suivante. Cette solution répond également à CVE -2021-45046.

Note

Les scripts d'action bootstrap pour Amazon EMR ont été mis à jour le 7 septembre 2022 pour inclure des corrections de bogues progressives et des améliorations pour Oozie. Si vous utilisez Oozie, vous devez appliquer la solution d'action Amazon EMR bootstrap mise à jour décrite dans la section suivante.

Amazon EMR sur EKS

Si vous utilisez [Amazon EKS avec EMR la](#) configuration par défaut, vous n'êtes pas concerné par le problème décrit dans CVE -2021-44228 et vous n'êtes pas obligé d'appliquer la solution décrite dans la section. [Solution d'action Amazon EMR Bootstrap pour Log4j CVE -2021-44228 et -2021-45046 CVE](#) Pour Amazon EMR onEKS, le EMR runtime Amazon pour Spark utilise Apache Log4j version 1.2.17. Lorsque vous utilisez Amazon EMR sur, EKS vous ne devez pas modifier le paramètre par défaut `log4j.appenders` du composant `surlog`.

Solution d'action Amazon EMR Bootstrap pour Log4j CVE -2021-44228 et -2021-45046 CVE

Cette solution fournit une action Amazon EMR bootstrap qui doit être appliquée à vos EMR clusters Amazon. Pour chaque EMR version d'Amazon, vous trouverez ci-dessous un lien vers un script d'action bootstrap. Pour appliquer cette action d'amorçage, vous devez suivre les étapes suivantes :

1. Copiez le script correspondant à votre EMR version d'Amazon dans un compartiment S3 local de votre Compte AWS. Assurez-vous que vous utilisez un script bootstrap spécifique à votre EMR version d'Amazon.
2. Configurez une action bootstrap pour que vos EMR clusters exécutent le script copié dans votre compartiment S3 conformément aux instructions décrites dans [EMR la documentation](#). Si d'autres actions d'amorçage sont configurées pour vos EMR clusters, assurez-vous que ce script est configuré comme le premier script d'action de démarrage à exécuter.
3. Mettez fin aux EMR clusters existants et lancez de nouveaux clusters à l'aide du script d'action bootstrap. AWS recommande de tester les scripts bootstrap dans votre environnement de test et de valider vos applications avant de les appliquer à votre environnement de production. Si vous n'utilisez pas la dernière version pour une version EMR mineure (par exemple, 6.3.0), vous devez utiliser la dernière révision (par exemple, 6.3.1), puis appliquer la solution décrite ci-dessus.

CVE-2021-44228 & CVE -2021-45046 - Scripts Bootstrap pour les versions Amazon EMR

Numéro EMR de version Amazon	Emplacement du script	Date de sortie du script
6.5.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.5.0-v2.sh</code>	24 mars 2022
6.4.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.4.0-v2.sh</code>	24 mars 2022
6.3.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.3.1-v2.sh</code>	24 mars 2022
6.2.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.2.1-v2.sh</code>	24 mars 2022
6.1.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.1.1-v2.sh</code>	14 décembre 2021
6.0.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.0.1-v2.sh</code>	14 décembre 2021
5,34,0	<code>s3://elasticmapreduce/ bootstrap-actions/</code>	12 décembre 2021

Numéro EMR de version Amazon	Emplacement du script	Date de sortie du script
	<code>log4j/patch-log4j-emr-5.34.0-v2.sh</code>	
5,33.1	<code>s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.33.1-v2.sh</code>	12 décembre 2021
5,32.1	<code>s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.32.1-v2.sh</code>	13 décembre 2021
5,31.1	<code>s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.31.1-v2.sh</code>	13 décembre 2021
5,30,2	<code>s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.30.2-v2.sh</code>	14 décembre 2021
5,29,0	<code>s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.29.0-v2.sh</code>	14 décembre 2021
5,28.1	<code>s3://elasticmapreduce/bootstrap-actions/log4j/patch-log4j-emr-5.28.1-v2.sh</code>	15 décembre 2021

Numéro EMR de version Amazon	Emplacement du script	Date de sortie du script
5,27.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.27.1-v2.sh</code>	15 décembre 2021
5,26,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.26.0-v2.sh</code>	15 décembre 2021
5,25,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.25.0-v2.sh</code>	15 décembre 2021
5.24.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.24.1-v2.sh</code>	15 décembre 2021
5.23.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.23.1-v2.sh</code>	15 décembre 2021
5,22,0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.22.0-v2.sh</code>	15 décembre 2021

Numéro EMR de version Amazon	Emplacement du script	Date de sortie du script
5.21.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.21.2-v2.sh</code>	15 décembre 2021
5.20.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.20.1-v2.sh</code>	15 décembre 2021
5.19.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.19.1-v2.sh</code>	15 décembre 2021
5.18.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.18.1-v2.sh</code>	15 décembre 2021
5,17.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.17.2-v2.sh</code>	15 décembre 2021
5.16.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.16.1-v2.sh</code>	15 décembre 2021

Numéro EMR de version Amazon	Emplacement du script	Date de sortie du script
5.15.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.15.1-v2.sh</code>	15 décembre 2021
5.14.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.14.2-v2.sh</code>	15 décembre 2021
5.13.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.13.1-v2.sh</code>	15 décembre 2021
5.12.3	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.12.3-v2.sh</code>	15 décembre 2021
5.11.4	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.11.4-v2.sh</code>	15 décembre 2021
5.10.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.10.1-v2.sh</code>	15 décembre 2021

Numéro EMR de version Amazon	Emplacement du script	Date de sortie du script
5.9.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.9.1-v2.sh</code>	15 décembre 2021
5.8.3	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.8.3-v2.sh</code>	15 décembre 2021
5.7.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.7.1-v2.sh</code>	15 décembre 2021

EMRversion de sortie	Dernière révision en date de décembre 2021
6.3.0	6.3.1
6.2.0	6.2.1
6.1.0	6.1.1
6.0.0	6.0.1
5,33,0	5,33.1
5,32,0	5,32.1
5,31,0	5,31.1
5.30.0 ou 5.30.1	5,30,2
5,28,0	5,28.1

EMRversion de sortie	Dernière révision en date de décembre 2021
5,27,0	5,27.1
5,24,0	5.24.1
5.23.0	5.23.1
5.21.0 ou 5.21.1	5.21.2
5,20,0	5.20.1
5,19,0	5.19.1
5,18,0	5.18.1
5.17.0 ou 5.17.1	5,17.2
5.16.0	5.16.1
5.15.0	5.15.1
5.14.0 ou 5.14.1	5.14.2
5.13.0	5.13.1
5.12.0, 5.12.1, 5.12.2	5.12.3
5.11.0, 5.11.1, 5.11.2, 5.11.3	5.11.4
5.9.0	5.9.1
5,8,0, 5,8.1, 5,8.2	5.8.3
5.7.0	5.7.1

Questions fréquentes (FAQ)

- Les EMR versions antérieures à EMR 5 sont-elles concernées par CVE -2021-44228 ?

Non. EMR les versions antérieures à la EMR version 5 utilisent des versions de Log4j antérieures à 2.0.

- Cette solution répond-elle à CVE -2021-45046 ?

Oui, cette solution répond également à CVE-2021-45046.

- La solution gère-t-elle les applications personnalisées que j'installe sur mes EMR clusters ?

Le script bootstrap met uniquement à jour JAR les fichiers installés par EMR. Si vous installez et exécutez des applications et des JAR fichiers personnalisés sur vos EMR clusters par le biais d'actions de démarrage, d'étapes soumises à vos clusters, en utilisant Amazon Linux AMI personnalisé ou par le biais de tout autre mécanisme, contactez votre fournisseur d'applications pour déterminer si vos applications personnalisées sont affectées par le CVE -2021- 44228 et pour déterminer une solution appropriée.

- Comment gérer les [images docker personnalisées](#) avec EMR on ? EKS

Si vous ajoutez des applications personnalisées à Amazon EMR EKS en utilisant des [images docker personnalisées](#) ou si vous soumettez des offres d'emploi à Amazon EMR sur des fichiers de candidature EKS with personnalisés, contactez le fournisseur de l'application pour déterminer si vos applications personnalisées sont concernées par CVE -2021-44228 et pour déterminer une solution appropriée.

- Comment fonctionne le script bootstrap pour atténuer le problème décrit dans CVE -2021-44228 et -2021-45046 ? CVE

Le script bootstrap met à jour les instructions de EMR démarrage en ajoutant un nouvel ensemble d'instructions. Ces nouvelles instructions suppriment les fichiers JndiLookup de classe utilisés via Log4j par tous les frameworks open source installés par EMR. Cela fait suite à la [recommandation publiée par Apache](#) pour résoudre les problèmes liés à Log4j.

- Existe-t-il une mise à jour EMR qui utilise les versions 2.17.1 ou supérieures de Log4j ?

EMR5 versions jusqu'à la version 5.34 et EMR 6 versions jusqu'à la version 6.5 utilisent des versions plus anciennes de frameworks open source incompatibles avec les dernières versions de Log4j. Si vous continuez à utiliser ces versions, nous vous recommandons d'appliquer l'action bootstrap pour atténuer les problèmes décrits dans le CVEs. Après les versions EMR 5 5.34 et EMR 6 version 6.5, les applications qui utilisent Log4j 1.x et Log4j 2.x seront mises à niveau pour utiliser Log4j 1.2.17 (ou version ultérieure) et Log4j 2.17.1 (ou version ultérieure) respectivement,

et ne nécessiteront pas d'utiliser les actions d'amorçage fournies ci-dessus pour atténuer les problèmes. CVE

- Les EMR versions sont-elles affectées par CVE -2021-45105 ?

Les applications installées par Amazon EMR avec EMR les configurations par défaut ne sont pas affectées par CVE -2021-45105. Parmi les applications installées par AmazonEMR, seul Apache Hive utilise Apache Log4j avec des [recherches contextuelles](#), et il n'utilise pas de disposition de modèle autre que celle par défaut de manière à permettre le traitement de données d'entrée inappropriées.

- Amazon est-il EMR concerné par l'une des CVE divulgations suivantes ?

Le tableau suivant contient une liste de ceux CVEs qui sont liés à Log4j et indique si chacun d'entre eux a un impact CVE sur Amazon. EMR Les informations de ce tableau ne s'appliquent que lorsque les applications sont installées par Amazon EMR à l'aide des configurations par défaut.

CVE	Répercussions EMR	Remarques
CVE-2_23302	Non	Amazon EMR ne configure pas Log4j JMSSink
CVE-2_23305	Non	Amazon EMR ne configure pas Log4j JDBCAppender
CVE-2_23307	Non	Amazon EMR ne configure pas Log4j Chainsaw
CVE-2020-9493	Non	Amazon EMR ne configure pas Log4j Chainsaw
CVE-2021-44832	Non	Amazon EMR ne configure pas Log4j JDBCAppender avec une JNDI chaîne de connexion
CVE-2021-4104	Non	Amazon EMR n'utilise pas Log4j JMSAppender

CVE	Répercussions EMR	Remarques
CVE-2020-9488	Non	Les applications installées par Amazon n'utilisent EMR pas Log4j SMTPAppender
CVE-2019-17571	Non	Amazon EMR bloque l'accès public aux clusters et ne les lance pas SocketServer
CVE-2019-17531	Non	Nous vous recommandons de passer à la dernière EMR version d'Amazon. Amazon EMR 5.33.0 et versions ultérieures utilisent jackson-databind 2.6.7.4 ou version ultérieure, et jackson-databind 2.10.0 ou version ultérieure. EMR Ces versions de jackson-databind ne sont pas affectées par le. CVE

EMR Archive des notes de publication sur Amazon

Les notes de publication pour toutes les EMR versions d'Amazon sont disponibles ci-dessous. Pour obtenir des informations complètes sur chaque version, consultez [Versions publiées par Amazon EMR 6.x](#), [Versions publiées par Amazon EMR 5.x](#) et [Versions publiées par Amazon EMR 4.x](#).

Pour recevoir des mises à jour lorsqu'une nouvelle EMR version d'Amazon est disponible, abonnez-vous au [RSSflux des notes de EMR publication d'Amazon](#).

Version 6.14.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.14.0 d'Amazon. Les modifications ont été apportées à la version 6.13.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.14.0](#).

Nouvelles fonctionnalités

- Amazon EMR 6.14.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-2, Flink 1.17.1, Iceberg 1.3.1, and Trino 422.
- [Amazon EMR Managed Scaling](#) est désormais disponible dans la région ap-southeast-3 Asie-Pacifique (Jakarta) pour les clusters que vous créez avec Amazon EMR 6.14.0 ou version ultérieure.

Modifications, améliorations et problèmes résolus

- La version 6.14.0 optimise la gestion des journaux avec Amazon exécuté EMR sur Amazon. EC2 C'est pourquoi vous constaterez peut-être une légère réduction des coûts de stockage pour les journaux de votre cluster.
- La version 6.14.0 améliore le flux de travail de dimensionnement afin de prendre en compte les différentes instances principales dont la taille varie considérablement par rapport à leurs volumes AmazonEBS. Cette amélioration s'applique uniquement aux nœuds principaux ; les opérations de réduction des nœuds de tâches ne sont pas affectées.
- La version 6.14.0 améliore la façon dont Amazon EMR interagit avec les applications open source telles que Apache Hadoop YARN ResourceManager and HDFS NameNode Cette amélioration réduit le risque de retards opérationnels liés à la mise à l'échelle du cluster et atténue les échecs de démarrage dus à des problèmes de connectivité avec les applications open source.
- La version 6.14.0 optimise l'installation des applications au lancement du cluster. Cela améliore les temps de démarrage du cluster pour certaines combinaisons d'EMRapplications Amazon.
- La version 6.14.0 résout un problème selon lequel les opérations de réduction du cluster pouvaient se bloquer lorsqu'un cluster exécuté dans un VPC domaine personnalisé redémarre un nœud principal ou un nœud de tâche.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne) , Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023.101,0	4,1,4327	17 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)
2,0.2023 906.0	4,1,4322	11 septembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

Version 6.13.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.13.0 d'Amazon. Les modifications ont été apportées à la version 6.12.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.13.0](#).

Nouvelles fonctionnalités

- Amazon EMR 6.13.0 supports Apache Spark 3.4.1, Apache Spark RAPIDS 23.06.0-amzn-1, CUDA Toolkit 11.8.0, and JupyterHub 1.5.0.

Modifications, améliorations et problèmes résolus

- La version 6.13.0 améliore le démon de gestion des EMR journaux Amazon afin de garantir que tous les journaux sont chargés à une cadence régulière vers Amazon S3 lorsqu'une commande de fin de cluster est émise. Cela permet de résilier plus rapidement les clusters.
- La version 6.13.0 améliore les fonctionnalités de gestion des EMR journaux d'Amazon afin de garantir le téléchargement cohérent et rapide de tous les fichiers journaux sur Amazon S3. Cela profite particulièrement aux EMR clusters de longue durée.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfor

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			t), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelea Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelea Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)
2,0.2023 4,1,4322 906.0		4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

Version 6.12.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.12.0 d'Amazon. Les modifications ont été apportées à la version 6.11.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.12.0](#).

Nouvelles fonctionnalités

- Amazon EMR 6.12.0 supports Apache Spark 3.4.0, Apache Spark RAPIDS 23.06.0-amzn-0, CUDA 11.8.0, Apache Hudi 0.13.1-amzn-0, Apache Iceberg 1.3.0-amzn-0, Trino 414, and PrestoDB 0.281.
- EMR Les versions 6.12.0 et supérieures d'Amazon prennent en charge LDAP l'intégration avec Apache Livy, Apache Hive via HiveServer 2 (HS2), Trino, Presto et Hue. Vous pouvez également installer Apache Spark et Apache Hadoop sur un EMR cluster utilisant la version 6.12.0 ou une version ultérieure et les configurer pour les utiliser. LDAP Pour plus d'informations, consultez [Utiliser Active Directory ou LDAP des serveurs pour l'authentification auprès d'Amazon EMR](#).

Modifications, améliorations et problèmes résolus

- EMR Les versions 6.12.0 et supérieures d'Amazon fournissent un support d'exécution Java 11 pour Flink. Pour de plus amples informations, veuillez consulter [Configurer Flink pour qu'il fonctionne avec Java 11](#).
- La version 6.12.0 ajoute un nouveau mécanisme de nouvelle tentative au flux de travail de dimensionnement des EMR clusters pour les clusters qui exécutent Presto ou Trino. Cette amélioration réduit le risque que le redimensionnement du cluster soit bloqué indéfiniment en raison de l'échec d'une seule opération de redimensionnement. Cela améliore également l'utilisation du cluster, car celui-ci augmente et diminue la capacité plus rapidement.
- La version 6.12.0 corrige un problème à cause duquel les opérations de réduction de la taille du cluster peuvent se bloquer lorsqu'un nœud principal en cours de mise hors service appropriée tombe en panne pour une raison quelconque avant d'être complètement mis hors service.
- La version 6.12.0 améliore la logique de réduction du cluster afin que celui-ci ne tente pas de réduire les nœuds principaux en dessous du paramètre de facteur de HDFS réplication défini pour le cluster. Cela répond à vos exigences en matière de redondance des données et réduit le risque de blocage d'une opération de dimensionnement.
- La version 6.12.0 améliore les performances et l'efficacité du service de surveillance de l'état pour Amazon EMR en augmentant la vitesse à laquelle il enregistre les changements d'état des

instances. Cette amélioration réduit le risque de dégradation des performances des nœuds de cluster qui exécutent plusieurs outils clients personnalisés ou des applications tierces.

- La version 6.12.0 améliore les performances du démon de gestion des journaux sur le cluster pour Amazon. EMR Par conséquent, les risques de dégradation des performances sont moindres avec les EMR clusters qui exécutent des étapes avec une grande simultanéité.
- Avec la EMR version 6.12.0 d'Amazon, le démon de gestion des journaux a été mis à niveau pour identifier tous les journaux utilisés activement avec des descripteurs de fichiers ouverts sur le stockage d'instance local, ainsi que les processus associés. Cette mise à niveau garantit qu'Amazon supprime EMR correctement les fichiers et récupère de l'espace de stockage une fois les journaux archivés dans Amazon S3.
- La version 6.12.0 inclut une amélioration du démon de gestion des journaux qui supprime les répertoires d'étapes vides et inutilisés dans le système de fichiers du cluster local. Un trop grand nombre de répertoires vides peut dégrader les performances des EMR démons Amazon et entraîner une surutilisation des disques.
- La version 6.12.0 permet la rotation des journaux pour les journaux de YARN Timeline Server. Cela permet de minimiser les scénarios de surutilisation des disques, en particulier pour les clusters de longue durée.
- La taille du volume racine par défaut est passée à 15 Go dans Amazon EMR 6.10.0 et versions ultérieures. Les versions antérieures ont une taille de volume racine par défaut de 10 Go.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023-0822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre)

Version 6.11.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.11.1 d'Amazon. Les modifications ont été apportées à la version 6.11.0. Pour plus d'informations sur le calendrier de publication, consultez le [Journal des modifications 6.11.1](#).

Modifications, améliorations et problèmes résolus

- En raison d'un conflit de verrouillage, un nœud peut se retrouver bloqué s'il est ajouté ou supprimé en même temps qu'il est mis hors service. Par conséquent, le gestionnaire de ressources Hadoop (YARN) ne répond plus et affecte tous les conteneurs entrants et en cours d'exécution.
- Cette version inclut une modification qui permet aux clusters à haute disponibilité de se remettre d'un état défaillant après le redémarrage.
- Cette version inclut des correctifs de sécurité pour Hue etHBase.
- Cette version résout un problème selon lequel les clusters exécutant des charges de travail sur Spark avec Amazon EMR peuvent recevoir silencieusement des résultats incorrects avec `contains`, `startsWith`, `endsWith`, et `like`. Ce problème se produit lorsque vous utilisez les expressions sur des champs partitionnés contenant des métadonnées sur le serveur Amazon EMR Hive3 Metastore (`hive.metastore`). HMS
- Cette version corrige un problème de régulation du côté de Glue lorsqu'il n'existe aucune fonction définie par l'utilisateur (`glue.udf`). UDF
- Cette version corrige un problème qui supprime les journaux des conteneurs par le service d'agrégation des journaux des nœuds avant que le lanceur de journaux ne puisse les envoyer vers S3 en cas de YARN mise hors service.
- Cette version corrige un problème lié aux métriques du FairShare planificateur lorsque Node Label est activé pour Hadoop.
- Cette version corrige un problème qui affectait les performances de Spark lorsque vous définissez une valeur `true` par défaut pour la configuration `spark.yarn.heterogeneousExecutors.enabled` dans `spark-defaults.conf`.
- Cette version corrige un problème lié à l'échec de la lecture des données de shuffle par Reduce Task. Ce problème provoquait des échecs de requêtes Hive avec une erreur de mémoire corrompue.
- Cette version ajoute un nouveau mécanisme de nouvelle tentative au flux de travail de dimensionnement des EMR clusters pour les clusters qui exécutent Presto ou Trino. Cette amélioration réduit le risque que le redimensionnement du cluster soit bloqué indéfiniment.

en raison de l'échec d'une seule opération de redimensionnement. Cela améliore également l'utilisation du cluster, car celui-ci augmente et diminue la capacité plus rapidement.

- Cette version améliore la logique de réduction du cluster afin que votre cluster ne tente pas de réduire les nœuds principaux en dessous du paramètre de facteur de HDFS réplication défini pour le cluster. Cela répond à vos exigences en matière de redondance des données et réduit le risque de blocage d'une opération de dimensionnement.
- Le démon de gestion des journaux a été mis à niveau pour identifier tous les journaux en cours d'utilisation avec des descripteurs de fichiers ouverts sur le stockage d'instance local, ainsi que les processus associés. Cette mise à niveau garantit qu'Amazon supprime EMR correctement les fichiers et récupère de l'espace de stockage une fois les journaux archivés dans Amazon S3.
- Cette version inclut une amélioration du démon de gestion des journaux qui supprime les répertoires d'étapes vides et inutilisés dans le système de fichiers du cluster local. Un trop grand nombre de répertoires vides peut dégrader les performances des EMR démons Amazon et entraîner une surutilisation des disques.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Paci

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			<p>Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie-Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie-Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)</p>

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia), Canada Ouest (Calgary)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre), Israël (Tel Aviv), AWS GovCloud (États-Unis ouest), AWS GovCloud

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(États-Unis est), Chine (Pékin), Chine (Ningxia)
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023-0822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

Version 6.11.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.11.0 d'Amazon. Les modifications ont été apportées à la version 6.10.0. Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Nouvelles fonctionnalités

- Amazon EMR 6.11.0 prend en charge Apache Spark 3.3.2-amzn-0, Apache Spark RAPIDS 23.02.0-amzn-0, 11.8.0, Apache Hudi 0.13.0-amzn-0, Apache Iceberg CUDA 1.2.0-amzn-0, Trino 410-amzn-0 et PrestoDB 0.279-amzn-0.

Modifications, améliorations et problèmes résolus

- Avec Amazon EMR 6.11.0, le connecteur DynamoDB a été mis à niveau vers la version 5.0.0. La version 5.0.0 utilise AWS SDK for Java 2.x. Les versions précédentes utilisaient la version AWS SDK for Java 1.x. À la suite de cette mise à niveau, nous vous conseillons vivement de tester votre code avant d'utiliser le connecteur DynamoDB avec Amazon 6.11. EMR
- Lorsque le connecteur DynamoDB pour EMR Amazon 6.11.0 appelle le service DynamoDB, il utilise la valeur de région que vous fournissez pour la propriété `dynamodb.endpoint`. Nous vous recommandons de configurer également `dynamodb.region` lorsque vous utilisez `dynamodb.endpoint`, et que les deux propriétés ciblent la même Région AWS. Si vous l'utilisez `dynamodb.endpoint` et que vous ne le configurez pas `dynamodb.region`, le connecteur DynamoDB pour EMR Amazon 6.11.0 renverra une exception de région non valide et tentera de réconcilier vos Région AWS informations à partir du service de métadonnées d'instance EC2 Amazon (). IMDS Si le connecteur ne parvient pas à récupérer la région IMDS, la valeur par défaut est des États-Unis (Virginie du Nord) (`us-east-1`). L'erreur suivante illustre l'exception de région non valide que vous pourriez obtenir si vous ne configurez pas correctement la `dynamodb.region` propriété : `error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region.` pour plus d'informations sur les classes concernées par la AWS SDK for Java mise à niveau vers la version 2.x, consultez le commit [Upgrade AWS SDK for Java from 1.x to 2.x \(#175\)](#) dans le GitHub dépôt du connecteur Amazon - EMR DynamoDB.
- Cette version corrige un problème où les données des colonnes deviennent NULL lorsque vous utilisez Delta Lake pour stocker les données de la table Delta dans Amazon S3 après l'opération de changement de nom de colonne. Pour plus d'informations sur cette fonctionnalité expérimentale dans Delta Lake, consultez [Opération de changement de nom de colonne](#) dans le guide de l'utilisateur Delta Lake.
- La version 6.11.0 résout un problème qui peut survenir lorsque vous créez un nœud périphérique en répliquant l'un des nœuds primaires à partir d'un cluster comportant plusieurs nœuds primaires. Le nœud périphérique répliqué peut retarder les opérations de réduction d'échelle ou entraîner une


utilisation élevée de la mémoire sur les nœuds primaires. Pour plus d'informations sur la création d'un nœud périphérique pour communiquer avec votre EMR cluster, consultez la section [Edge Node Creator](#) dans le `aws-samples` référentiel sur GitHub.

- La version 6.11.0 améliore le processus d'automatisation EMR utilisé par Amazon pour remonter les volumes EBS Amazon sur une instance après un redémarrage.
- La version 6.11.0 corrige un problème qui entraînait des écarts intermittents dans les métriques Hadoop publiées par Amazon sur Amazon. EMR CloudWatch
- La version 6.11.0 résout un problème lié aux EMR clusters où une mise à jour du fichier de YARN configuration contenant la liste d'exclusion des nœuds du cluster est interrompue en raison d'une surutilisation du disque. La mise à jour incomplète entrave les futures opérations de réduction de la taille du cluster. Cette version garantit que votre cluster reste sain et que les opérations de dimensionnement fonctionnent comme prévu.
- La taille du volume racine par défaut est passée à 15 Go dans Amazon EMR 6.10.0 et versions ultérieures. Les versions antérieures ont une taille de volume racine par défaut de 10 Go.
- Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans activation de la mise à l'échelle gérée.

Dans les EMR versions 6.11.0 et supérieures d'Amazon ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` est définie sur `false` in `yarn-site.xml` pour résoudre ce problème.

Bien que le correctif corrige les problèmes introduits par YARN -9608, il peut entraîner l'échec des tâches Hive en raison de la perte de données par shuffle sur les clusters sur lesquels le dimensionnement géré est activé. Nous avons atténué ce risque dans cette version en définissant également `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` pour les charges de travail Hive. Cette configuration n'est disponible qu'avec les EMR versions 6.11.0 et supérieures d'Amazon.

- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

 Note

Cette version ne reçoit plus de AMI mises à jour automatiques puisqu'elle a été remplacée par une autre version de correctif. La version du correctif est indiquée par le numéro qui suit la deuxième décimale (6 . 8 . **1**). Pour savoir si vous utilisez la dernière version du correctif, consultez les versions disponibles dans le [guide des versions](#), ou consultez le menu déroulant des EMRversions d'Amazon lorsque vous créez un cluster dans la console, ou utilisez l'[list-release-labels](#) CLI action [ListReleaseLabels](#) API ou. Pour être tenu au courant des nouvelles versions, abonnez-vous au RSS fil d'actualité sur le [site Quoi de neuf ?](#) page.

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre), Israël (Tel Aviv)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre)

Version 6.10.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.10.0 d'Amazon. Les modifications ont été apportées à la version 6.9.0. Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Nouvelles fonctionnalités

- Amazon EMR 6.10.0 prend en charge Apache Spark 3.3.1, Apache Spark RAPIDS 22.12.0, CUDA 11.8.0, Apache Hudi 0.12.2-amzn-0, Apache Iceberg 1.1.0-amzn-0, Trino 403 et PrestoDB 0.278.1.
- Amazon EMR 6.10.0 inclut un connecteur Trino-Hudi natif qui fournit un accès en lecture aux données des tables Hudi. Vous pouvez activer le connecteur avec `trino-cli --catalog hudi` et le configurer en fonction de vos besoins avec `trino-connector-hudi`. L'intégration native avec Amazon EMR signifie que vous n'avez plus besoin de l'utiliser `trino-connector-hive` pour interroger les tables Hudi. Pour obtenir la liste des configurations prises en charge avec le nouveau connecteur, consultez la page [Connecteur Hudi](#) dans la documentation de Trino.
- EMR Les versions 6.10.0 et supérieures d'Amazon prennent en charge l'intégration d'Apache Zeppelin avec Apache Flink. Pour plus d'informations, consultez [Travailler avec les jobs Flink de Zeppelin dans Amazon EMR](#).

Problèmes connus

- Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans activation de la mise à l'échelle gérée.

Pour contourner ce problème dans Amazon EMR 6.10.0, vous pouvez définir la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications to false` in `yarn-site.xml` Dans les EMR versions 6.11.0 et supérieures d'Amazon, ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la configuration est définie `false` par défaut pour résoudre ce problème.

Modifications, améliorations et problèmes résolus

- Amazon EMR 6.10.0 supprime la dépendance à l'[intégration d'Amazon Redshift minimal-json.jar pour Apache Spark](#) et ajoute automatiquement les fichiers jar liés à Spark-Redshift requis au chemin de classe de l'exécuteur pour Spark :, et. `spark-redshift.jar` `spark-avro.jar` `RedshiftJDBC.jar`
- La version 6.10.0 améliore le démon de gestion des journaux sur le cluster afin de surveiller des dossiers de journaux supplémentaires dans votre cluster. EMR Cette amélioration permet de minimiser les scénarios de surutilisation des disques.
- La version 6.10.0 redémarre automatiquement le démon de gestion des journaux du cluster lorsqu'il s'arrête. Cette amélioration réduit le risque que les nœuds apparaissent défectueux en raison d'une surutilisation du disque.
- Amazon EMR 6.10.0 prend en charge les points de terminaison régionaux pour EMRFS le mappage des utilisateurs.
- La taille du volume racine par défaut est passée à 15 Go dans Amazon EMR 6.10.0 et versions ultérieures. Les versions antérieures ont une taille de volume racine par défaut de 10 Go.
- La version 6.10.0 corrige un problème qui provoquait le blocage des tâches Spark lorsque tous les exécuteurs Spark restants se trouvaient sur un hôte en cours de mise hors service avec le gestionnaire de ressources. YARN
- Avec Amazon EMR 6.6.0 à 6.9.x, les INSERT requêtes comportant une partition dynamique et une clause ORDER BY ou SORT BY auront toujours deux réducteurs. Ce problème est dû à la OSS modification [HIVE-20703](#), qui soumet l'optimisation des partitions de tri dynamique à une décision basée sur les coûts. Si votre charge de travail ne nécessite pas le tri des partitions dynamiques, nous vous recommandons de définir la propriété `hive.optimize.sort.dynamic.partition.threshold` sur -1 pour désactiver la nouvelle fonctionnalité et obtenir le nombre de réducteurs correctement calculé. Ce problème est résolu dans OSS Hive dans le cadre de la version [HIVE-22269](#) et dans Amazon 6.10.0. EMR
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

Note

Cette version ne reçoit plus de AMI mises à jour automatiques puisqu'elle a été remplacée par une autre version de correctif. La version du correctif est indiquée par le numéro qui

suit la deuxième décimale (6.8.1). Pour savoir si vous utilisez la dernière version du correctif, consultez les versions disponibles dans le [guide des versions](#), ou consultez le menu déroulant des EMRversions d'Amazon lorsque vous créez un cluster dans la console, ou utilisez l'[list-release-labels](#) CLI action [ListReleaseLabels](#) API ou. Pour être tenu au courant des nouvelles versions, abonnez-vous au RSS fil d'actualité sur le [site Quoi de neuf ?](#) page.

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Zurich), Europe (Milan), Europe (Espagne), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad) Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient () UAE

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong Kong), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong Kong), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023-07.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong Kong), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE)

Version 6.9.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.9.0 d'Amazon. Les modifications concernent la EMR version 6.8.0 d'Amazon. Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Nouvelles fonctions

- La EMR version 6.9.0 d'Amazon prend en charge Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 et Tez 0.10.2.
- La EMR version 6.9.0 d'Amazon inclut une nouvelle application open source, [Delta Lake](#) la 2.1.0.
- L'intégration Amazon Redshift pour Apache Spark est incluse dans les EMR versions 6.9.0 et ultérieures d'Amazon. Auparavant un outil open-source, l'intégration native est un connecteur Spark que vous pouvez utiliser pour créer des applications Apache Spark capables de lire et d'écrire des données sur Amazon Redshift et Amazon Redshift sans serveur. Pour de plus amples informations, veuillez consulter [Utilisation de l'intégration d'Amazon Redshift pour Apache Spark avec Amazon EMR](#).
- La EMR version 6.9.0 d'Amazon ajoute la prise en charge de l'archivage des journaux dans Amazon S3 lors de la réduction du cluster. Auparavant, vous pouviez uniquement archiver les fichiers journaux sur Amazon S3 lors de la résiliation du cluster. Cette nouvelle fonctionnalité garantit que les fichiers journaux générés sur le cluster sont conservés sur Amazon S3 même après la résiliation du nœud. Pour plus d'informations, consultez [Configuration de la journalisation et du débogage de cluster](#).
- Pour prendre en charge les requêtes de longue durée, Trino inclut désormais un mécanisme d'exécution tolérant aux pannes. L'exécution tolérante aux pannes atténue les échecs des requêtes en réessayant les requêtes qui ont échoué ou les tâches correspondantes. Pour de plus amples informations, veuillez consulter [Exécution tolérante aux pannes dans Trino](#).
- Vous pouvez utiliser Apache Flink sur Amazon EMR pour unifier BATCH et STREAM traiter les tables Apache Hive ou les métadonnées de n'importe quelle source de table Flink telle qu'Iceberg, Kinesis ou Kafka. Vous pouvez spécifier le catalogue de données AWS Glue comme métastore pour Flink à l'aide de AWS Management Console, AWS CLI ou Amazon. EMR API Pour de plus amples informations, veuillez consulter [Configuration de Flink sur Amazon EMR](#).
- Vous pouvez désormais spécifier des rôles d'exécution AWS Identity and Access Management (IAM) et un contrôle d'accès AWS Lake Formation basé pour les requêtes Apache Spark, Apache Hive et Presto EMR sur Amazon sur des EC2 clusters avec Amazon SageMaker Studio. Pour plus d'informations, consultez [Configurer les rôles d'exécution pour Amazon EMR Steps](#).

Problèmes connus

- Pour la EMR version 6.9.0 d'Amazon, Trino ne fonctionne pas sur les clusters activés pour Apache Ranger. Si vous devez utiliser Trino avec Ranger, contactez [AWS Support](#).

- Si vous utilisez l'intégration Amazon Redshift à Apache Spark et que vous disposez d'un champ de type heure, timetz, horodatage ou timestamptz avec une précision de l'ordre de la microseconde au format Parquet, le connecteur arrondit les valeurs temporelles à la milliseconde la plus proche. Pour contourner le problème, utilisez le paramètre `unload_s3_format` de format de téléchargement du texte.
- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

- Les connexions aux EMR clusters Amazon depuis Amazon SageMaker Studio peuvent échouer par intermittence avec un code de réponse 403 Forbidden. Cette erreur se produit lorsque la configuration du IAM rôle sur le cluster prend plus de 60 secondes. Pour contourner le problème, vous pouvez installer un EMR correctif Amazon pour activer les nouvelles tentatives et augmenter le délai d'expiration à un minimum de 300 secondes. Suivez les étapes ci-dessous pour appliquer l'action d'amorçage lorsque vous lancez votre cluster.

1. Téléchargez le script et RPM les fichiers bootstrap depuis l'Amazon S3 URIs suivant.

```
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/replace-rpms.sh
s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/gcsc/emr-secret-agent-1.18.0-SNAPSHOT20221121212949.noarch.rpm
```

2. Chargez les fichiers de l'étape précédente sur un compartiment Amazon S3 dont vous êtes propriétaire. Le compartiment doit se trouver à l' Région AWS endroit où vous prévoyez de lancer le cluster.
3. Incluez l'action d'amorçage suivante lorsque vous lancez votre EMR cluster. Remplacez *bootstrap_URI* and *RPM_URI* avec le correspondant URIs provenant d'Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- Avec les EMR versions 5.36.0 et 6.6.0 à 6.9.0 d'Amazon, SecretAgent les composants du RecordServer service peuvent subir une perte de données de journal en raison d'une configuration incorrecte du modèle de nom de fichier dans les propriétés Log4j2. En cas de configuration incorrecte, les composants ne génèrent qu'un seul fichier journal par jour. Lorsque la stratégie de rotation est appliquée, elle remplace le fichier existant au lieu de générer un nouveau fichier journal comme prévu. Pour contourner le problème, utilisez une action d'amorçage pour générer des journaux toutes les heures et ajoutez un nombre entier auto-incrémenté dans le nom du fichier pour gérer la rotation.

Pour les versions Amazon EMR 6.6.0 à 6.9.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Pour Amazon EMR 5.36.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Apache Flink fournit des FileSystem connecteurs natifs S3 FileSystem et Hadoop, qui permettent aux applications de créer FileSink et d'écrire les données dans Amazon S3. Cela FileSink échoue avec l'une des deux exceptions suivantes.

```
java.lang.UnsupportedOperationException: Recoverable writers on Hadoop are only supported for HDFS
```

```
Caused by: java.lang.NoSuchMethodError: org.apache.hadoop.io.retry.RetryPolicies.retryOtherThanRemoteAndSaslException(Lorg/
```

```

apache/hadoop/io/retry/RetryPolicy;Ljava/util/Map;)Lorg/apache/hadoop/io/retry/
RetryPolicy;

                                at
  org.apache.hadoop.yarn.client.RMProxy.createRetryPolicy(RMProxy.java:302) ~[hadoop-
yarn-common-3.3.3-amzn-0.jar:?]

```

Pour contourner le problème, vous pouvez installer un EMR correctif Amazon, qui corrige le problème ci-dessus dans Flink. Suivez les étapes suivantes pour appliquer l'action d'amorçage lors du lancement de votre cluster.

1. Téléchargez le fichier flink-rpm dans votre compartiment Amazon S3. Votre RPM voie l'est `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Téléchargez le script et les RPM fichiers bootstrap depuis Amazon S3 en utilisant ce qui suit URL. Remplacez `regionName` par l' Région AWS endroit où vous prévoyez de lancer le cluster.

```

s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-
rpms.sh

```

3. Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Dans Amazon EMR 6.8.0 et 6.9.0, cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans le dimensionnement géré activé.

Avec [Amazon EMR 6.10.0](#), il existe une solution à ce problème qui consiste à définir la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications to in. false` dans `yarn-site.xml`. Dans les EMR versions 6.11.0 et supérieures d'Amazon, ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la configuration est définie `false` par défaut pour résoudre ce problème.

Modifications, améliorations et problèmes résolus

- Pour les EMR versions 6.9.0 et ultérieures d'Amazon, tous les composants installés par Amazon EMR qui utilisent les bibliothèques Log4j utilisent Log4j version 2.17.1 ou ultérieure.
- Lorsque vous utilisez le connecteur DynamoDB avec Spark on EMR Amazon versions 6.6.0, 6.7.0 et 6.8.0, toutes les lectures de votre table renvoient un résultat vide, même si le split d'entrée fait référence à des données non vides. La EMR version 6.9.0 d'Amazon résout ce problème.

- Amazon EMR 6.9.0 ajoute une prise en charge limitée du contrôle d'accès basé sur Lake Formation avec Apache Hudi lors de la lecture de données à l'aide de Spark. SQL Le support concerne les SELECT requêtes utilisant Spark SQL et se limite au contrôle d'accès au niveau des colonnes. Pour plus d'informations, consultez [Hudi et Lake Formation](#).
- Lorsque vous utilisez Amazon EMR 6.9.0 pour créer un cluster Hadoop avec les [étiquettes de nœuds](#) activées, les [YARNmétriques API](#) renvoient des informations agrégées sur toutes les partitions, au lieu de la partition par défaut. Pour plus d'informations, consultez [YARN-11414](#).
- Avec la EMR version 6.9.0 d'Amazon, nous avons mis à jour Trino vers la version 398, qui utilise Java 17. La précédente version prise en charge de Trino pour Amazon EMR 6.8.0 était Trino 388 fonctionnant sous Java 11. Pour plus d'informations sur cette modification, consultez [Mises à jour de Trino vers Java 17](#) sur le blog de Trino.
- Cette version corrige un problème d'inadéquation des séquences temporelles entre Apache BigTop et Amazon EMR lors de la séquence de démarrage EC2 du cluster. Ce décalage se produit lorsqu'un système tente d'effectuer deux ou plusieurs opérations en même temps au lieu de les effectuer dans le bon ordre. Par conséquent, certaines configurations de cluster ont connu des délais de démarrage des instances et des temps de démarrage des clusters plus lents.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

Note

Cette version ne reçoit plus de AMI mises à jour automatiques puisqu'elle a été remplacée par une autre version de correctif. La version du correctif est indiquée par le numéro qui suit la deuxième décimale (6 . 8 . **1**). Pour savoir si vous utilisez la dernière version du correctif, consultez les versions disponibles dans le [guide des versions](#), ou consultez le menu déroulant des EMRversions d'Amazon lorsque vous créez un cluster dans la console, ou utilisez l'[list-release-labels](#) CLI action [ListReleaseLabels](#) API. Pour être tenu au courant des nouvelles versions, abonnez-vous au RSS fil d'actualité sur le [site Quoi de neuf ?](#) page.

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Melbourne), Israël (Tel Aviv), Canada Ouest (Calgary)
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre), Israël (Tel Aviv)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central), Israël (Tel Aviv)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelea Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 307.0	4,14,305	15 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 207.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202210.1	4,14,301	12 janvier 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 103.3	4,14.296	5 décembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

Version 6.8.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.8.0 d'Amazon. Les modifications ont été apportées à la version 6.7.0.

Nouvelles fonctions

- La fonctionnalité Amazon EMR Steps prend désormais en charge le point de terminaison JDBC et/ou les ODBC clients Apache Livy. Pour plus d'informations, consultez [Configurer les rôles d'exécution pour Amazon EMR Steps](#).
- La EMR version 6.8.0 d'Amazon est fournie avec la HBase version 2.4.12 d'Apache. Avec cette HBase version, vous pouvez à la fois archiver et supprimer vos HBase tables. Le processus d'archivage Amazon S3 renomme tous les fichiers de table dans le répertoire d'archive. Ce processus peut être long et coûteux. Vous pouvez désormais ignorer le processus d'archivage et supprimer rapidement des tables volumineuses. Pour de plus amples informations, veuillez consulter [Utilisation de la HBase coque](#).

Problèmes connus

- Hadoop 3.3.3 a introduit une modification dans YARN ([YARN-9608](#)) qui maintient les nœuds sur lesquels les conteneurs s'exécutaient dans un état de mise hors service jusqu'à ce que l'application soit terminée. Cette modification permet de s'assurer que les données locales telles que les données réorganisées ne sont pas perdues et que vous n'avez pas besoin de réexécuter la tâche. Dans Amazon EMR 6.8.0 et 6.9.0, cette approche peut également entraîner une sous-utilisation des ressources sur les clusters avec ou sans le dimensionnement géré activé.

Avec [Amazon EMR 6.10.0](#), il existe une solution à ce problème qui consiste à définir la valeur de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `in. false` dans `yarn-site.xml`. Dans les EMR versions 6.11.0 et supérieures d'Amazon, ainsi que dans les versions 6.8.1, 6.9.1 et 6.10.1, la configuration est définie `false` par défaut pour résoudre ce problème.

Modifications, améliorations et problèmes résolus

- Lorsque les EMR versions 6.5.0, 6.6.0 ou 6.7.0 d'Amazon lisaient les tables Apache Phoenix via le shell Apache Spark, Amazon EMR a produit un `NoSuchMethodError`. La EMR version 6.8.0 d'Amazon résout ce problème.
- La EMR version 6.8.0 d'Amazon est fournie avec [Apache Hudi](#) 0.11.1 ; toutefois, les clusters Amazon EMR 6.8.0 sont également compatibles avec l'open source de Hudi 0.12.0. `hudi-spark3.3-bundle_2.12`

- La EMR version 6.8.0 d'Amazon est fournie avec Apache Spark 3.3.0. Cette version de Spark utilise Apache Log4j 2 et le fichier `log4j2.properties` pour configurer Log4j dans les processus Spark. Si vous utilisez Spark dans le cluster ou si vous créez des EMR clusters avec des paramètres de configuration personnalisés, et que vous souhaitez passer à la EMR version 6.8.0 d'Amazon, vous devez migrer vers la nouvelle classification de `spark-log4j2` configuration et le nouveau format de clé pour Apache Log4j 2. Pour de plus amples informations, veuillez consulter [Migration d'Apache Log4j 1.x vers Log4j 2.x](#).
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

Note

Cette version ne reçoit plus de AMI mises à jour automatiques puisqu'elle a été remplacée par une autre version de correctif. La version du correctif est indiquée par le numéro qui suit la deuxième décimale (6.8.1). Pour savoir si vous utilisez la dernière version du correctif, consultez les versions disponibles dans le [guide des versions](#), ou consultez le menu déroulant des EMRversions d'Amazon lorsque vous créez un cluster dans la console, ou utilisez l'[list-release-labels](#) CLI action [ListReleaseLabels](#) API ou. Pour être tenu au courant des nouvelles versions, abonnez-vous au RSS fil d'actualité sur le [site Quoi de neuf ?](#) page.

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14.348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich), Asie-Pacifique (Melbourne), Israël (Tel Aviv)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Asie-Pacifique (Melbourne), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Central)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.202307.0	4,14,305	15 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 207.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023119.1	4,14,301	3 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202210.1	4,14,301	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 103.3	4,14.296	5 décembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 004.0	4,14.294	2 novembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 912.1	4,14,291	6 septembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

Problèmes connus

- Lorsque vous utilisez le connecteur DynamoDB avec Spark on EMR Amazon versions 6.6.0, 6.7.0 et 6.8.0, toutes les lectures de votre table renvoient un résultat vide, même si le split d'entrée fait référence à des données non vides. Cela est dû au fait que Spark 3.2.0 définit `spark.hadoopRDD.ignoreEmptySplits` sur `true` par défaut. Pour contourner le problème,

définissez explicitement `spark.hadoopRDD.ignoreEmptySplits` sur `false`. La EMR version 6.9.0 d'Amazon résout ce problème.

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

- Avec les EMR versions 5.36.0 et 6.6.0 à 6.9.0 d'Amazon, `SecretAgent` les composants du `RecordServer` service peuvent subir une perte de données de journal en raison d'une configuration incorrecte du modèle de nom de fichier dans les propriétés `Log4j2`. En cas de configuration incorrecte, les composants ne génèrent qu'un seul fichier journal par jour. Lorsque la stratégie de rotation est appliquée, elle remplace le fichier existant au lieu de générer un nouveau fichier journal comme prévu. Pour contourner le problème, utilisez une action d'amorçage pour générer des journaux toutes les heures et ajoutez un nombre entier auto-incrémenté dans le nom du fichier pour gérer la rotation.

Pour les versions Amazon EMR 6.6.0 à 6.9.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Pour Amazon EMR 5.36.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Pour plus d'informations sur le calendrier de publication, consultez le [journal des modifications](#).

Version 6.7.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.7.0 d'Amazon. Les modifications ont été apportées à la version 6.6.0.

Date de parution initiale : 15 juillet 2022

Nouvelles fonctions

- Amazon prend EMR désormais en charge Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 et Trino 0.378.
- Prend en charge les contrôles d'accès basés sur les IAM rôles et les formations lacustres avec des EMR étapes (Spark, Hive) pour Amazon EMR sur les clusters. EC2
- Prend en charge les instructions de définition de données Apache Spark sur les clusters compatibles avec Apache Ranger. Cela inclut désormais la prise en charge des applications Trino lisant et écrivant des métadonnées Apache Hive sur des clusters compatibles avec Apache Ranger. Pour plus d'informations, consultez [Activer la gouvernance fédérée à l'aide de Trino et Apache Ranger sur Amazon](#). EMR
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio),

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia), Asie Pacifique (Hyderabad), Moyen-Orient (UAE), Europe (Espagne), Europe (Zurich)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023.0.926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 307.0	4,14,305	15 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023-07.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023119.1	4,14,301	3 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202210.1	4,14,301	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 103.3	4,14.296	5 décembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022004.0	4,14.294	2 novembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2022 912.1	4,14,291	7 octobre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0,2022 719,0	4,14,287	10 août 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 606.1	4,14,281	15 juillet 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

Problèmes connus

- Lorsque les EMR versions 6.5.0, 6.6.0 ou 6.7.0 d'Amazon lisent les tables Apache Phoenix via le shell Apache Spark, cela se produit `NoSuchMethodError` car Amazon EMR utilise une valeur incorrecte. `Hbase.compat.version` La EMR version 6.8.0 d'Amazon résout ce problème.
- Lorsque vous utilisez le connecteur DynamoDB avec Spark on EMR Amazon versions 6.6.0, 6.7.0 et 6.8.0, toutes les lectures de votre table renvoient un résultat vide, même si le split

d'entrée fait référence à des données non vides. Cela est dû au fait que Spark 3.2.0 définit `spark.hadoopRDD.ignoreEmptySplits` sur `true` par défaut. Pour contourner le problème, définissez explicitement `spark.hadoopRDD.ignoreEmptySplits` sur `false`. La EMR version 6.9.0 d'Amazon résout ce problème.

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

- Avec les EMR versions 5.36.0 et 6.6.0 à 6.9.0 d'Amazon, `SecretAgent` les composants du `RecordServer` service peuvent subir une perte de données de journal en raison d'une configuration incorrecte du modèle de nom de fichier dans les propriétés `Log4j2`. En cas de configuration incorrecte, les composants ne génèrent qu'un seul fichier journal par jour. Lorsque la stratégie de rotation est appliquée, elle remplace le fichier existant au lieu de générer un nouveau fichier journal comme prévu. Pour contourner le problème, utilisez une action d'amorçage pour générer des journaux toutes les heures et ajoutez un nombre entier auto-incrémenté dans le nom du fichier pour gérer la rotation.

Pour les versions Amazon EMR 6.6.0 à 6.9.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Pour Amazon EMR 5.36.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- Il `GetClusterSessionCredentials` API n'est pas pris en charge avec les clusters qui s'exécutent sur Amazon EMR 6.7 ou une version antérieure.

Version 6.6.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.6.0 d'Amazon. Les modifications ont été apportées à la version 6.5.0.

Date de parution initiale : 9 mai 2022

Date de mise à jour de la documentation : 15 juin 2022

Nouvelles fonctions

- Amazon EMR 6.6 prend désormais en charge Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 et PrestoDB 0.267.
- Lorsque vous lancez un cluster avec le dernier correctif d'Amazon EMR 5.36 ou supérieur, 6.6 ou supérieur, ou 7.0 ou supérieur, Amazon EMR utilise la dernière version d'Amazon Linux 2023 ou Amazon Linux 2 pour Amazon EMR AMI par défaut. Pour plus d'informations, consultez [Utiliser le système Amazon Linux par défaut AMI pour Amazon EMR](#).

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 709.1	4,14348	23 juillet 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
			(Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 223.0	4,14.336	8 mars 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 131.0	4,14.336	14 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 124.0	4,14.336	7 février 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2024 109.0	4,1,4334	24 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 218,0	4,14,330	2 janvier 2024	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 206,0	4,14,330	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 116,0	4,14,328	11 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 101,0	4,1,4327	16 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023020.1	4,1,4326	7 novembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 012.1	4,1,4326	26 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023.0.926,0	4,1,4322	19 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Singapour), Asie-Pacifique (Asie Pacifique) (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (centre), (États-Unis ouest), AWS GovCloud (États-Unis est), Chine AWS GovCloud (Pékin), Chine (Ningxia)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 906.0	4,1,4322	4 octobre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 822.0	4,1,4322	30 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 808.0	4,14,320	24 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 727,0	4,14,320	14 août 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Francfort), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 719,0	4,14,320	02/08/2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Europe (Stockholm), Europe (Milan), Europe (Espagne), Europe (Francfort), Europe (Zurich), Europe (Irlande), Europe (Londres), Europe (Paris), Asie-Pacifique (Hong Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Hyderabad), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Asie-Pacifique (Jakarta), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn), Moyen-Orient (UAE), Canada (Centre)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 628,0	4,14,318	12 juillet 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 612,0	4,1,4314	23 juin 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 504.1	4,14.313	16 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 418,0	4,14.311	3 mai 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023 404.1	4,14.311	18 avril 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0.2023 404.0	4,14.311	10 avril 2023	USA Est (Virginie du Nord), Europe (Paris)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2023 320,0	4,14,309	30 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023 307.0	4,14,305	15 mars 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2023-07.0	4,14,304	22 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.2023119.1	4,14,301	3 février 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2.0.202210.1	4,14,301	22 décembre 2023	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsReleaseLabel (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 103.3	4,14.296	5 décembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022004.0	4,14.294	2 novembre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2022 912.1	4,14,291	7 octobre 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)
2,0,2022 805.0	4,14,287	30 août 2022	us-west-1

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0,2022 719,0	4,14.287	10 août 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 426.0	4,14,281	10 juin 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

OsRelease Label (Version Amazon Linux)	Version du noyau Amazon Linux	Date de disponibilité	Régions prises en charge
2,0.2022 406.1	4,14,275	2 mai 2022	USA Est (Virginie du Nord), USA Est (Ohio), USA Ouest (Californie du Nord), USA Ouest (Oregon), Canada (Centre), Europe (Stockholm), Europe (Irlande), Europe (Londres), Europe (Paris), Europe (Francfort), Europe (Milan), Asie-Pacifique (Hong-Kong), Asie-Pacifique (Mumbai), Asie-Pacifique (Jakarta), Asie-Pacifique (Tokyo), Asie-Pacifique (Séoul), Asie-Pacifique (Osaka), Asie-Pacifique (Singapour), Asie-Pacifique (Sydney), Afrique (Le Cap), Amérique du Sud (São Paulo), Moyen-Orient (Bahreïn)

- [Avec Amazon EMR 6.6 et versions ultérieures, les applications qui utilisent Log4j 1.x et Log4j 2.x sont mises à niveau pour utiliser Log4j 1.2.17 \(ou version ultérieure\) et Log4j 2.17.1 \(ou version ultérieure\) respectivement, et ne nécessitent pas d'utiliser les actions d'amorçage fournies pour atténuer les problèmes.](#) CVE
- [Dimensionnement géré] Optimisation du dimensionnement géré par les données Spark Shuffle - Pour EMR les versions 5.34.0 et ultérieures d'Amazon, ainsi que pour EMR les versions 6.4.0 et ultérieures, le dimensionnement géré prend désormais en compte les données Spark shuffle (données que Spark redistribue entre les partitions pour effectuer des opérations spécifiques).

Pour plus d'informations sur les opérations de shuffle, consultez la section [Utilisation du dimensionnement EMR géré EMR dans Amazon](#) dans le Guide de EMRgestion Amazon et le Guide de [programmation Spark](#).

- À partir d'Amazon EMR 5.32.0 et 6.5.0, le dimensionnement dynamique de l'exécuteur pour Apache Spark est activé par défaut. Pour activer ou désactiver cette fonctionnalité, vous pouvez utiliser le paramètre de configuration `spark.yarn.heterogeneousExecutors.enabled`.

Modifications, améliorations et problèmes résolus

- Amazon EMR réduit le temps de démarrage des clusters de 80 secondes en moyenne pour les clusters qui utilisent l'AMloption EMR par défaut et installent uniquement des applications courantes, telles qu'Apache Hadoop, Apache Spark et Apache Hive.

Problèmes connus

- Lorsque les EMR versions 6.5.0, 6.6.0 ou 6.7.0 d'Amazon lisent les tables Apache Phoenix via le shell Apache Spark, cela se produit `NoSuchMethodError` car Amazon EMR utilise une valeur incorrecte. `Hbase.compat.version` La EMR version 6.8.0 d'Amazon résout ce problème.
- Lorsque vous utilisez le connecteur DynamoDB avec Spark on EMR Amazon versions 6.6.0, 6.7.0 et 6.8.0, toutes les lectures de votre table renvoient un résultat vide, même si le split d'entrée fait référence à des données non vides. Cela est dû au fait que Spark 3.2.0 définit `spark.hadoopRDD.ignoreEmptySplits` sur `true` par défaut. Pour contourner le problème, définissez explicitement `spark.hadoopRDD.ignoreEmptySplits` sur `false`. La EMR version 6.9.0 d'Amazon résout ce problème.
- Sur les clusters de longue durée de Trino, Amazon EMR 6.6.0 active les paramètres de journalisation de Garbage Collection dans le fichier `jvm.config` de Trino afin d'obtenir de meilleures informations à partir des journaux de Garbage Collection. Cette modification ajoute de nombreux journaux du récupérateur de mémoire au fichier `launcher.log` (`/var/log/trino/launcher.log`). Si vous utilisez des clusters Trino dans Amazon EMR 6.6.0, vous risquez de rencontrer des nœuds à court d'espace disque après quelques jours d'exécution du cluster en raison des journaux ajoutés.

La solution à ce problème consiste à exécuter le script ci-dessous en tant qu'action Bootstrap afin de désactiver les paramètres de journalisation de Garbage Collection dans `jvm.config` lors de la création ou du clonage du cluster pour Amazon 6.6.0. EMR

```
#!/bin/bash
```

```
set -ex
PRESTO_PUPPET_DIR='/var/aws/emr/bigtop-deploy/puppet/modules/trino'
sudo bash -c "sed -i '/-Xlog/d' ${PRESTO_PUPPET_DIR}/templates/jvm.config"
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration

`spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

- Avec les EMR versions 5.36.0 et 6.6.0 à 6.9.0 d'Amazon, `SecretAgent` les composants du `RecordServer` service peuvent subir une perte de données de journal en raison d'une configuration incorrecte du modèle de nom de fichier dans les propriétés `Log4j2`. En cas de configuration incorrecte, les composants ne génèrent qu'un seul fichier journal par jour. Lorsque la stratégie de rotation est appliquée, elle remplace le fichier existant au lieu de générer un nouveau fichier journal comme prévu. Pour contourner le problème, utilisez une action d'amorçage pour générer des journaux toutes les heures et ajoutez un nombre entier auto-incrémenté dans le nom du fichier pour gérer la rotation.

Pour les versions Amazon EMR 6.6.0 à 6.9.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Pour Amazon EMR 5.36.0, utilisez l'action bootstrap suivante lorsque vous lancez un cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Version 5.35.0

Ceci est la note de mise à jour de la EMR version 5.35.0 d'Amazon.

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.35.0 d'Amazon. Les modifications ont été apportées à la version 5.34.0.

Date de parution initiale : 30 mars 2022

Nouvelles fonctions

- Les applications Amazon EMR version 5.35 qui utilisent Log4j 1.x et Log4j 2.x sont mises à niveau pour utiliser Log4j 1.2.17 (ou version ultérieure) et Log4j 2.17.1 (ou version ultérieure) respectivement, et ne nécessitent pas d'actions d'amorçage pour atténuer les problèmes rencontrés dans les versions précédentes. CVE Consultez [Approche d'atténuation CVE-2021-44228](#).

Modifications, améliorations et problèmes résolus

Modifications apportées à Flink

Type de modification	Description
Mises à niveau	<ul style="list-style-type: none">• Mise à jour de la version de flink vers la version 1.14.2.• log4j a été mis à jour vers la version 2.17.1.

Modifications apportées à Hadoop

Type de modification	Description
Rétroportages open source Hadoop depuis la version 5.34.0 EMR	<ul style="list-style-type: none"> • YARN-10438 : Gérer la valeur nulle container Id en C lientRMService # getContainerReport () • YARN-7266 : Les fils du gestionnaire d'événements de Timeline Server sont verrouillés • YARN-10438 : ATS 1.5 ne démarre pas si RollingLevelDb les fichiers sont corrompus ou manquants • HADOOP-13500 : Synchronisation de l'itération de l'objet des propriétés de configuration • YARN-10651 : CapacityScheduler s'est écrasé avec NPE in. AbstractYarnScheduler updateNodeResource() • HDFS-12221 : Remplacer les xerces dans XmlEditsVisitor • HDFS-16410 : Analyse XML non sécurisée OfflineEditsXmlLoader
Modifications et corrections apportées à Hadoop	<ul style="list-style-type: none"> • Tomcat est utilisé dans KMS et HTTPFS est mis à niveau vers la version 8.5.75 • Dans la FileSystemOptimizedCommitter version 2, le marqueur de réussite était écrit dans le chemin commitJob de sortie défini lors de la création du commit. Étant donné que commitJob les chemins de sortie au niveau des tâches peuvent différer, le chemin a été corrigé pour utiliser celui défini dans les fichiers manifestes. Pour les tâches Hive, cela signifie que le marqueur de réussite est correctement écrit lors de l'exécution

Type de modification	Description
	d'opérations telles que la partition dynamique ou UNIONALL.

Modifications apportées à Hive

Type de modification	Description
Hive est passé à la version open source 2.3.9 , y compris ces correctifs JIRA	<ul style="list-style-type: none"> • HIVE-17155 : findConfFile () dans le HiveConf fichier .java présente quelques problèmes avec le chemin de configuration • HIVE-24797 : Désactiver la validation des valeurs par défaut lors de l'analyse des schémas Avro • HIVE-21563 : Améliorez les getEmptyTable performances de Table# en désactivant Once registerAllFunctions • HIVE-18147 : Les tests peuvent échouer avec java.net. BindException: Adresse déjà utilisée • HIVE-24608 : Revenir à get_table dans HMS le client pour Hive 2.3.x • HIVE-21200 : Vectorisation : colonne de date lançant java.lang. UnsupportedOperationException pour parquet • HIVE-19228 : Supprime l'utilisation de commons-httpclient 3.x
Rétroportages open source Hive depuis la version 5.34.0 EMR	<ul style="list-style-type: none"> • HIVE-19990 : La requête avec un intervalle littéral dans la condition de jointure échoue • HIVE-25824 : Mise à niveau de la branche 2.3 vers log4j 2.17.0

Type de modification	Description
	<ul style="list-style-type: none"> • TEZ-4062 : La planification des tentatives spéculatives doit être abandonnée une fois la tâche terminée • TEZ-4108 : NullPointerException lors d'une exécution spéculative, condition de course • TEZ-3918 : La configuration de <code>tez.task.log.level</code> ne fonctionne pas
Mises à niveau et correctifs apportés à Hive	<ul style="list-style-type: none"> • Mise à niveau de la version de Log4j vers 2.17.1 • Mise à niveau de ORC la version vers la version 1.4.3 • Correction d'un blocage dû à un fil de pénalité ShuffleScheduler
Nouvelles fonctionnalités	<ul style="list-style-type: none"> • Ajout d'une fonctionnalité permettant d'imprimer Hive Query dans les journaux AM. Par défaut, l'option est désactivée. Drapeau/Conf : <code>tez.am.emr.print.hive.query.in.log</code> État (par défaut) :FALSE.

Modifications apportées à Oozie

Type de modification	Description
Rétroportages open source d'Oozie depuis la version 5.34.0 EMR	<ul style="list-style-type: none"> • OOZIE-3652 : Le lanceur Oozie devrait réessayer de répertorier les répertoires lorsque cela se produit NoSuchFileException

Modifications apportées à Pig

Type de modification	Description
Mises à niveau	<ul style="list-style-type: none">log4j a été mis à jour vers la version 1.2.17.

Problèmes connus

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Version 5.34.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.34.0 d'Amazon. Les modifications ont été apportées à la version 5.33.1.

Date de parution initiale : 20 janvier 2022

Date de mise à niveau : 21 mars 2022

Nouvelles fonctions

- [Dimensionnement géré] Optimisation du dimensionnement géré par les données Spark Shuffle - Pour EMR les versions 5.34.0 et ultérieures d'Amazon, ainsi que pour EMR les versions 6.4.0 et ultérieures, le dimensionnement géré prend désormais en compte les données Spark shuffle (données que Spark redistribue entre les partitions pour effectuer des opérations spécifiques). Pour plus d'informations sur les opérations de shuffle, consultez la section [Utilisation du dimensionnement EMR géré EMR dans Amazon](#) dans le Guide de EMRgestion Amazon et le Guide de [programmation Spark](#).
- [Hudi] Améliorations visant à simplifier la configuration de Hudi. Le contrôle de simultanéité optimiste a été désactivé par défaut.

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Auparavant, le redémarrage manuel du gestionnaire de ressources sur un cluster multi-maîtres obligeait les EMR démons Amazon du cluster, tels que Zookeeper, à recharger tous les nœuds précédemment désaffectés ou perdus dans le fichier znode de Zookeeper. Cela a entraîné le dépassement des limites par défaut dans certaines situations. Amazon supprime EMR désormais les enregistrements de nœuds mis hors service ou perdus datant de plus d'une heure du fichier Zookeeper et les limites internes ont été augmentées.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.

- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Mise à niveau de Zeppelin vers la version 0.10.0.
- Livy Fix – mise à niveau vers la version 0.7.1
- Amélioration des performances de Spark : les exécuteurs hétérogènes sont désactivés lorsque certaines valeurs de configuration de Spark sont remplacées dans la version 5.34.0. EMR
- Le Web HDFS et le serveur HTTPFS sont désactivés par défaut. Vous pouvez réactiver le Web à HDFS l'aide de la configuration Hadoop, `dfs.webhdfs.enabled` Le serveur HttpFS peut être démarré en utilisant `sudo systemctl start hadoop-httpfs`.

Problèmes connus

- La fonctionnalité Amazon EMR Notebooks utilisée avec l'usurpation d'identité d'un utilisateur de Livy ne fonctionne pas car le protocole HTTPFS est désactivé par défaut. Dans ce cas, le EMR bloc-notes ne peut pas se connecter au cluster sur lequel l'emprunt d'identité Livy est activé. La solution consiste à démarrer le serveur HTTPFS avant de connecter le EMR bloc-notes au cluster à l'aide de `sudo systemctl start hadoop-httpfs`
- Les requêtes Hue ne fonctionnent pas dans Amazon EMR 6.4.0 car le serveur Apache Hadoop HTTPFS est désactivé par défaut. Pour utiliser Hue on Amazon EMR 6.4.0, démarrez manuellement le serveur HTTPFS sur le nœud EMR principal Amazon à l'aide `sudo systemctl start hadoop-httpfs` ou [utilisez une étape Amazon EMR](#).
- La fonctionnalité Amazon EMR Notebooks utilisée avec l'usurpation d'identité d'un utilisateur de Livy ne fonctionne pas car le protocole HTTPFS est désactivé par défaut. Dans ce cas, le EMR bloc-notes ne peut pas se connecter au cluster sur lequel l'emprunt d'identité Livy est activé. La solution consiste à démarrer le serveur HTTPFS avant de connecter le EMR bloc-notes au cluster à l'aide de `sudo systemctl start hadoop-httpfs`
- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement

les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :

- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Version 6.5.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.5.0 d'Amazon. Les modifications ont été apportées à la version 6.4.0.

Date de parution initiale : 20 janvier 2022

Date de mise à niveau : 21 mars 2022

Nouvelles fonctions

- [Dimensionnement géré] Optimisation du dimensionnement géré par les données Spark Shuffle - Pour EMR les versions 5.34.0 et ultérieures d'Amazon, ainsi que pour EMR les versions 6.4.0 et ultérieures, le dimensionnement géré prend désormais en compte les données Spark shuffle (données que Spark redistribue entre les partitions pour effectuer des opérations spécifiques). Pour plus d'informations sur les opérations de shuffle, consultez la section [Utilisation du dimensionnement EMR géré EMR dans Amazon](#) dans le Guide de EMRgestion Amazon et le Guide de [programmation Spark](#).
- À partir d'Amazon EMR 5.32.0 et 6.5.0, le dimensionnement dynamique de l'exécuteur pour Apache Spark est activé par défaut. Pour activer ou désactiver cette fonctionnalité, vous pouvez utiliser le paramètre de configuration `spark.yarn.heterogeneousExecutors.enabled`.

- Prise en charge du format de table ouvert Apache Iceberg pour les jeux de données analytiques volumineux.
- Support pour ranger-trino-plugin 2.0.1-amzn-1
- Prise en charge de toree 0.5.0

Modifications, améliorations et problèmes résolus

- La version EMR 6.5 d'Amazon prend désormais en charge Apache Iceberg 0.12.0 et fournit des améliorations d'exécution avec Amazon EMR Runtime pour Apache Spark, Amazon EMR Runtime pour Presto et Amazon EMR Runtime pour Apache Hive.
- [Apache Iceberg](#) est un format de table ouvert pour les grands ensembles de données dans Amazon S3. Il fournit des performances de requête rapides sur de grandes tables, des validations atomiques, des écritures simultanées et une évolution de table SQL compatible. Avec la EMR version 6.5, vous pouvez utiliser Apache Spark 3.1.2 avec le format de table Iceberg.
- Apache Hudi 0.9 ajoute Spark SQL DDL et le DML support. Cela vous permet de créer et de modifier des tables Hudi en utilisant uniquement SQL des instructions. Apache Hudi 0.9 inclut également des améliorations des performances côté requête et côté écriture.
- Amazon EMR Runtime pour Apache Hive améliore les performances d'Apache Hive sur Amazon S3 en supprimant les opérations de renommage lors des opérations intermédiaires, et améliore les performances des commandes metastore check (MSCK) utilisées pour réparer les tables.

Problèmes connus

- Lorsque les EMR versions 6.5.0, 6.6.0 ou 6.7.0 d'Amazon lisent les tables Apache Phoenix via le shell Apache Spark, cela se produit `NoSuchMethodError` car Amazon EMR utilise une valeur incorrecte. `Hbase.compat.version` La EMR version 6.8.0 d'Amazon résout ce problème.
- Les clusters de la solution groupée Hbase en haute disponibilité (HA) ne parviennent pas à se provisionner avec la taille de volume et le type d'instance par défaut. La solution à ce problème consiste à augmenter la taille du volume racine.
- Pour utiliser les actions Spark avec Apache Oozie, vous devez ajouter la configuration suivante à votre fichier Oozie `workflow.xml`. Sinon, plusieurs bibliothèques critiques telles que Hadoop et B EMRFS seront absentes du classpath des exécuteurs Spark lancés par Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-`. Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Version 6.4.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.4.0 d'Amazon. Les modifications ont été apportées à la version 6.3.0.

Date de parution initiale : 20 septembre 2021

Date de mise à niveau : 21 mars 2022

Applications prises en charge

- AWS SDK for Java version 1.12.31
- CloudWatch Version 2.2.0 de l'évier
- Connecteur DynamoDB version 4.16.0
- EMRFSversion 2.47.0
- Amazon EMR Goodies version 3.2.0
- Connecteur Amazon EMR Kinesis version 3.5.0

- Amazon EMR Record Server version 2.1.0
- Amazon EMR Scripts version 2.5.0
- Flink version 1.13.1
- Ganglia version 3.7.2
- AWS Client Glue Hive Metastore version 3.3.0
- Hadoop version 3.2.1-amzn-4
- HBaseversion 2.4.4-amzn-0
- HBase-operator-tools 1.1.0
- HCatalogversion 3.1.2-amzn-5
- Hive version 3.1.2-amzn-5
- Hudi version 0.8.0-amzn-0
- Hue version 4.9.0
- JDKVersion Java Corretto-8.302.08.1 (version 1.8.0_302-b08)
- JupyterHub version 1.4.1
- Livy version 0.7.1-incubating
- MXNetversion 1.8.0
- Oozie version 5.2.1
- Phoenix version 5.1.2
- Pig version 0.17.0
- Presto version 0.254.1-amzn-0
- Trino version 359
- Apache Ranger KMS (chiffrement transparent multi-maîtres) version 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.2.0
- SageMaker SDKVersion 1.4.1 de Spark
- Scala version 2.12.10 (machine virtuelle de serveur ouverte JDK 64 bits, Java 1.8.0_282)
- Spark version 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop version 1.4.7

- TensorFlow version 2.4.1
- tez version 0.9.2
- Zeppelin version 0.9.0
- Zookeeper version 3.5.7
- Connecteurs et pilotes : Connecteur DynamoDB 4.16.0

Nouvelles fonctionnalités

- [Dimensionnement géré] Optimisation du dimensionnement géré par les données Spark Shuffle - Pour EMR les versions 5.34.0 et ultérieures d'Amazon, ainsi que pour EMR les versions 6.4.0 et ultérieures, le dimensionnement géré prend désormais en compte les données Spark shuffle (données que Spark redistribue entre les partitions pour effectuer des opérations spécifiques). Pour plus d'informations sur les opérations de shuffle, consultez la section [Utilisation du dimensionnement EMR géré EMR dans Amazon](#) dans le Guide de EMRgestion Amazon et le Guide de [programmation Spark](#).
- Sur les EMR clusters Amazon compatibles avec Apache Ranger, vous pouvez utiliser Apache Spark SQL pour insérer des données dans les tables de métastore Apache Hive ou les mettre à jour à l'aide de, et. INSERT INTO INSERT OVERWRITE ALTER TABLE Lors de l'utilisation ALTER TABLE avec SparkSQL, l'emplacement d'une partition doit être le répertoire enfant d'un emplacement de table. Amazon EMR ne prend actuellement pas en charge l'insertion de données dans une partition où l'emplacement de la partition est différent de celui de la table.
- Presto SQL a été [renommé Trino](#).
- Hive : L'exécution de SELECT requêtes simples avec LIMIT clause est accélérée en arrêtant l'exécution de la requête dès que le nombre d'enregistrements mentionnés dans la LIMIT clause est récupéré. Les SELECT requêtes simples sont des requêtes qui ne contiennent pas de clause GROUP ORDER BY/by ou des requêtes qui n'ont pas d'étape de réduction. Par exemple, SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>.

Contrôles de simultanéité Hudi

- Hudi prend désormais en charge Optimistic Concurrency Control (OCC), qui peut être utilisé avec des opérations d'écriture telles que UPSERT et INSERT pour autoriser les modifications de plusieurs rédacteurs vers la même table Hudi. C'est au niveau du fichierOCC, donc deux validations (ou auteurs) peuvent écrire dans la même table, si leurs modifications ne sont pas contradictoires. Pour plus d'informations, consultez le [contrôle de simultanéité de Hudi](#).

- Zookeeper est installé sur les EMR clusters Amazon, qui peut être utilisé comme fournisseur de verrous pour OCC. Pour faciliter l'utilisation de cette fonctionnalité, les propriétés préconfigurées des EMR clusters Amazon sont les suivantes :

```
hoodie.write.lock.provider=org.apache.hudi.client.transaction.lock.ZookeeperBasedLockProvider
hoodie.write.lock.zookeeper.url=<EMR Zookeeper URL>
hoodie.write.lock.zookeeper.port=<EMR Zookeeper Port>
hoodie.write.lock.zookeeper.base_path=/hudi
```

Pour les activer OCC, vous devez configurer les propriétés suivantes soit avec leurs options de tâche Hudi, soit au niveau du cluster à l'aide des configurations Amazon : EMR API

```
hoodie.write.concurrency.mode=optimistic_concurrency_control
hoodie.cleaner.policy.failed.writes=LAZY (Performs cleaning of failed writes lazily
instead of inline with every write)
hoodie.write.lock.zookeeper.lock_key=<Key to uniquely identify the Hudi table> (Table
Name is a good option)
```

Hudi Monitoring : CloudWatch intégration d'Amazon pour générer des rapports sur Hudi Metrics

- Amazon EMR prend en charge la publication de Hudi Metrics sur Amazon CloudWatch. Elle est activée en définissant les configurations requises suivantes :

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- Les configurations Hudi facultatives que vous pouvez modifier sont les suivantes :

Paramètre	Description	Valeur
hoodie.metrics.cloudwatch.reporter.period.seconds	Fréquence (en secondes) à laquelle les statistiques doivent être communiquées à Amazon CloudWatch	La valeur par défaut est de 60 s, ce qui convient à la résolution d'une minute par défaut proposée par Amazon CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Préfixe à ajouter au nom de chaque métrique	La valeur par défaut est vide (pas de préfixe)

Paramètre	Description	Valeur
<code>hoodie.metrics.cloudwatch.namespace</code>	CloudWatch Espace de noms Amazon sous lequel les métriques sont publiées	La valeur par défaut est Hudi
<code>hoodie.metrics.cloudwatch.maxDatumsPerDemande</code>	Nombre maximum de références à inclure dans une demande adressée à Amazon CloudWatch	La valeur par défaut est 20, identique à la valeur CloudWatch par défaut d'Amazon

Support et améliorations des configurations Amazon EMR Hudi

- Les clients peuvent désormais tirer parti EMR de la fonctionnalité de configuration API et de reconfiguration pour configurer les configurations Hudi au niveau du cluster. Une nouvelle prise en charge de la configuration basée sur les fichiers a été introduite via `/etc/hudi/conf/hudi-defaults.conf`, à l'instar d'autres applications telles que Spark, Hive, etc. EMRconfigure quelques paramètres par défaut pour améliorer l'expérience utilisateur :
 - `hoodie.datasource.hive_sync.jdbcurl` est configuré sur le serveur Hive du cluster URL et n'a plus besoin d'être spécifié. Cela est particulièrement utile lorsque vous exécutez une tâche en mode cluster Spark, où vous deviez auparavant spécifier l'adresse IP EMR principale d'Amazon.
 - HBase des configurations spécifiques, utiles pour utiliser l'HBaseindex avec Hudi.
 - Configuration spécifique au fournisseur de verrous Zookeeper, comme indiqué dans la section Contrôle de simultanéité, qui facilite l'utilisation d'Optimistic Concurrency Control (). OCC
- Des modifications supplémentaires ont été introduites pour réduire le nombre de configurations à transmettre et pour en déduire automatiquement dans la mesure du possible :
 - Le mot clé `partitionBy` peut être utilisé pour spécifier la colonne de partition.
 - Lorsque vous activez Hive Sync, il n'est plus obligatoire de spécifier `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY`. Ces valeurs peuvent être déduites du nom de la table Hudi et du champ de partition.

— `KEYGENERATOR_CLASS_OPT_KEY` n'est pas obligatoire et peut être déduite de cas plus simples de `SimpleKeyGenerator` et `ComplexKeyGenerator`.

Mises en garde de Hudi

- Hudi ne prend pas en charge l'exécution vectorisée dans les tables Hive for Merge on Read (MoR) et Bootstrap. Par exemple, `count (*)` échoue avec la table en temps réel de Hudi lorsque `hive.vectorized.execution.enabled` est défini sur `true`. Comme solution de contournement, vous pouvez désactiver la lecture vectorisée en définissant `hive.vectorized.execution.enabled` sur `false`.
- La prise en charge des enregistreurs multiples n'est pas compatible avec la fonction d'amorçage de Hudi.
- Flink Streamer et Flink SQL sont des fonctionnalités expérimentales de cette version. Ces fonctionnalités ne sont pas recommandées pour les déploiements de production.

Modifications, améliorations et problèmes résolus

Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.

- Auparavant, le redémarrage manuel du gestionnaire de ressources sur un cluster multi-maîtres obligeait les EMR démons Amazon du cluster, tels que Zookeeper, à recharger tous les nœuds précédemment désaffectés ou perdus dans le fichier `znode` de Zookeeper. Cela a entraîné le dépassement des limites par défaut dans certaines situations. Amazon supprime EMR désormais les enregistrements de nœuds mis hors service ou perdus datant de plus d'une heure du fichier Zookeeper et les limites internes ont été augmentées.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.

- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Configuration d'un cluster pour résoudre les problèmes de performances du serveur Apache YARN Timeline versions 1 et 1.5

Les versions 1 et 1.5 d'Apache YARN Timeline Server peuvent entraîner des problèmes de performances avec de grands EMR clusters très actifs, en particulier avec `yarn.resourcemanager.system-metrics-publisher.enabled=true`, qui est le paramètre par défaut d'AmazonEMR. Un YARN Timeline Server v2 open source résout le problème de performance lié à l'évolutivité du YARN Timeline Server.

Les autres solutions à ce problème incluent :

- Configuration de `yarn.resourcemanager.system-metrics-publisher.enabled=false` dans le fichier `yarn-site.xml`.
- Activation du correctif pour ce problème lors de la création d'un cluster, comme décrit ci-dessous.

Les EMR versions Amazon suivantes contiennent un correctif pour ce problème de performance du serveur YARN Timeline.

EMR5,30.2, 5,31.1, 5,32.1, 5,33.1, 5,34.x, 6,0.1, 6.1.1, 6.2.1, 6.3.1, 6.4.x

Pour activer le correctif sur l'une des EMR versions Amazon spécifiées ci-dessus, définissez ces propriétés sur `true` dans un JSON fichier de configuration transmis à l'aide du [paramètre de aws emr create-cluster commande](#) : `--configurations file:///./configurations.json`.

Vous pouvez également activer le correctif à l'aide de [l'interface utilisateur de la console de reconfiguration](#).

Exemple du contenu du fichier configurations.json :

```
[
{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.resourcemanager.system-metrics-publisher.timeline-server-v1.enable-batch":
      "true",
    "yarn.resourcemanager.system-metrics-publisher.enabled": "true"
  },
  "Configurations": []
}
]
```

- Le Web HDFS et le serveur HTTPFS sont désactivés par défaut. Vous pouvez réactiver le Web à HDFS l'aide de la configuration Hadoop, `dfs.webhdfs.enabled`. Le serveur HttpFS peut être démarré en utilisant `sudo systemctl start hadoop-httpfs`.
- HTTPS est désormais activé par défaut pour les référentiels Amazon Linux. Si vous utilisez une VPCE politique Amazon S3 pour restreindre l'accès à des compartiments spécifiques, vous devez ajouter le nouveau compartiment Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à votre politique (remplacez-le par `$region` la région où se trouve le point de terminaison). Pour plus d'informations, consultez cette rubrique dans les forums de AWS discussion. [Annonce : Amazon Linux 2 permet désormais de l'utiliser HTTPS lors de la connexion à des référentiels de packages](#).
- Hive : les performances des requêtes d'écriture sont améliorées en permettant l'utilisation d'un répertoire temporaire activé HDFS pour la dernière tâche. Les données temporaires pour la tâche finale sont écrites sur Amazon S3 HDFS plutôt que sur Amazon S3 et les performances sont améliorées car les données sont déplacées depuis HDFS l'emplacement de la table finale (Amazon S3) plutôt qu'entre les appareils Amazon S3.
- Hive : amélioration du temps de compilation des requêtes jusqu'à 2,5 fois avec l'élimination des partitions du métastore Glue.
- Par défaut, lorsque les éléments intégrés UDFs sont transmis par Hive au serveur Hive Metastore, seul un sous-ensemble de ces éléments intégrés UDFs est transmis au Glue Metastore, car Glue ne prend en charge que des opérateurs d'expression limités. Si vous définissez `hive.glue.partition.pruning.client=true`, tout l'élimination des partitions se fait du côté

client. Si vous définissez `hive.glue.partition.pruning.server=true`, tout l'élimination des partitions se fait du côté serveur.

Problèmes connus

- Les requêtes Hue ne fonctionnent pas dans Amazon EMR 6.4.0 car le serveur Apache Hadoop HTTPFS est désactivé par défaut. Pour utiliser Hue on Amazon EMR 6.4.0, démarrez manuellement le serveur HTTPFS sur le nœud EMR principal Amazon à l'aide `sudo systemctl start hadoop-httpfs` ou [utilisez une étape Amazon EMR](#).
- La fonctionnalité Amazon EMR Notebooks utilisée avec l'usurpation d'identité d'un utilisateur de Livy ne fonctionne pas car le protocole HTTPFS est désactivé par défaut. Dans ce cas, le EMR bloc-notes ne peut pas se connecter au cluster sur lequel l'emprunt d'identité Livy est activé. La solution consiste à démarrer le serveur HTTPFS avant de connecter le EMR bloc-notes au cluster à l'aide de `sudo systemctl start hadoop-httpfs`
- Dans la EMR version 6.4.0 d'Amazon, Phoenix ne prend pas en charge le composant des connecteurs Phoenix.
- Pour utiliser les actions Spark avec Apache Oozie, vous devez ajouter la configuration suivante à votre fichier Oozie `workflow.xml`. Sinon, plusieurs bibliothèques critiques telles que Hadoop et B EMRFS seront absentes du classpath des exécuteurs Spark lancés par Oozie.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Version 5.32.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.32.0 d'Amazon. Les modifications ont été apportées à la version 5.31.0.

Date de parution initiale : 8 janvier 2021

Mises à niveau

- Mise à niveau du connecteur Amazon Glue vers la version 1.14.0
- Mise à niveau d'Amazon SageMaker Spark SDK vers la version 1.4.1
- Mise à niveau AWS SDK for Java vers la version 1.11.890
- Connecteur EMR DynamoDB version 4.16.0 mise à niveau
- Mise à niveau EMRFS vers la version 2.45.0
- Mise à niveau de EMR Log Analytics Metrics vers la version 1.18.0
- EMR MetricsAndEventsApiGateway Client mis à niveau vers la version 1.5.0
- Serveur d'EMRenregistrement mis à niveau vers la version 1.8.0
- Mise à niveau de EMR S3 Dist CP vers la version 2.17.0
- Agent EMR secret mis à niveau vers la version 1.7.0
- Mise à niveau de Flink vers la version 1.11.2
- Mise à niveau de Hadoop vers la version 2.10.1-amzn-0
- Mise à niveau de Hive vers la version 2.3.7-amzn-3
- Mise à niveau de Hue vers la version 4.8.0
- Mise à niveau de Mxnet vers la version 1.7.0
- Mise à niveau d'OpenCV vers la version 4.4.0
- Mise à niveau de Presto vers la version 0.240.1-amzn-0
- Mise à niveau de Spark vers la version 2.4.7-amzn-0

- Mise à niveau TensorFlow vers la version 2.3.1

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Versions de composants mises à niveau.
- Pour obtenir la liste des versions des composants, consultez la section [À propos des EMR versions d'Amazon](#) dans ce guide.

Nouvelles fonctionnalités

- À partir d'Amazon EMR 5.32.0 et 6.5.0, le dimensionnement dynamique de l'exécuteur pour Apache Spark est activé par défaut. Pour activer ou désactiver cette fonctionnalité, vous pouvez utiliser le paramètre de configuration `spark.yarn.heterogeneousExecutors.enabled`.
- État de prise en charge du service de métadonnées d'instance (IMDS) V2 : les composants Amazon EMR 5.23.1, 5.27.1 et 5.32 ou versions ultérieures sont utilisés pour tous les appels. IMDSv2 IMDS Pour les IMDS appels dans le code de votre application, vous pouvez utiliser les deux IMDSv1 et IMDSv2, ou les configurer IMDS pour les utiliser uniquement IMDSv2 pour renforcer la sécurité. Pour les autres EMR versions 5.x, la désactivation IMDSv1 entraîne l'échec du démarrage du cluster.
- À partir d'Amazon EMR 5.32.0, vous pouvez lancer un cluster qui s'intègre nativement à Apache Ranger. Apache Ranger est un cadre open source permettant d'activer, de surveiller et de gérer la sécurité globale des données sur la plateforme Hadoop. Pour plus d'informations, consultez [Apache Ranger](#). Grâce à l'intégration native, vous pouvez utiliser votre propre Apache Ranger pour appliquer un contrôle précis de l'accès aux données sur Amazon. EMR Consultez la section [Intégrer Amazon EMR à Apache Ranger](#) dans le guide de EMR mise à jour d'Amazon.
- La EMR version 5.32.0 d'Amazon est compatible avec Amazon EMR on. EKS Pour en savoir plus sur la prise en main EMR de EKS On, consultez [What is Amazon EMR on EKS](#).
- La EMR version 5.32.0 d'Amazon prend en charge Amazon EMR Studio (version préliminaire). Pour en savoir plus sur la prise en main de EMR Studio, consultez [Amazon EMR Studio \(version préliminaire\)](#).
- Politiques gérées délimitées : pour s'aligner sur les AWS meilleures pratiques, Amazon EMR a introduit des politiques gérées par EMR défaut définies dans la version v2 en remplacement des politiques qui seront obsolètes. Consultez les [politiques EMR gérées par Amazon](#).

Problèmes connus

- Pour les clusters de sous-réseaux privés Amazon EMR 6.3.0 et 6.2.0, vous ne pouvez pas accéder à l'interface utilisateur Web de Ganglia. Vous recevrez un message d'erreur « accès refusé (403) ». D'autres sites WebUIs, tels que Spark, Hue JupyterHub, Zeppelin, Livy et Tez, fonctionnent normalement. L'accès à l'interface utilisateur Web de Ganglia sur les clusters de sous-réseaux publics fonctionne également normalement. Pour résoudre ce problème, redémarrez le service `httpd` sur le nœud primaire avec `sudo systemctl restart httpd`. Ce problème est résolu dans Amazon EMR 6.4.0.

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux of Amazon 2 (), qui ont un paramètre `ulimit` inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre `ulimit` EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre `ulimit` inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre `ulimit` lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'`ulimit` du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un `ulimit` à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section `Service`.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer `InstanceController`

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un `ulimit` à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer `ulimit` du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

Important

EMR Les clusters qui exécutent Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) utilisent le comportement par défaut d'Amazon Linux et ne téléchargent ni n'installent automatiquement les mises à jour importantes et critiques du noyau nécessitant un redémarrage. Ce comportement est identique à celui des autres EC2 instances Amazon qui exécutent Amazon Linux par défautAMI. Si de nouvelles mises à jour logicielles Amazon Linux nécessitant un redémarrage (telles que le noyau et les CUDA mises à jour) sont disponibles après la EMR sortie d'une version d'Amazon, les instances de EMR cluster qui exécutent la version par défaut AMI ne téléchargent ni n'installent automatiquement ces mises à jour. NVIDIA Pour obtenir les mises à jour du noyau, vous pouvez [personnaliser votre Amazon EMR AMI](#) pour qu'il [utilise la dernière version d'Amazon Linux AMI](#).

- La prise en charge par console pour créer une configuration de sécurité spécifiant l'option d'intégration de AWS Ranger n'est actuellement pas prise en charge dans la GovCloud région. La configuration de sécurité peut être effectuée à l'aide duCLI. Consultez la section [Création EMR de la configuration de sécurité](#) dans le guide EMR de gestion Amazon.
- Lorsque AtRestEncryption HDFS le chiffrement est activé sur un cluster qui utilise Amazon EMR 5.31.0 ou 5.32.0, les requêtes Hive entraînent l'exception d'exécution suivante.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:
java.lang.RuntimeException: Hive Runtime Error while closing
operators: java.io.IOException: java.util.ServiceConfigurationError:
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Version 6.2.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.2.0 d'Amazon. Les modifications ont été apportées à la version 6.1.0.

Date de parution initiale : 9 décembre 2020

Dernière mise à jour : 4 octobre 2021

Applications prises en charge

- AWS SDK for Java version 1.11.828

- emr-record-server version 1.7.0
- Flink version 1.11.2
- Ganglia version 3.7.2
- Hadoop version 3.2.1-amzn-1
- HBaseversion 2.2.6-amzn-0
- HBase-operator-tools 1.0.0
- HCatalogversion 3.1.2-amzn-0
- Hive version 3.1.2-amzn-3
- Hudi version 0.6.0-amzn-1
- Hue version 4.8.0
- JupyterHub version 1.1.0
- Livy version 0.7.0
- MXNetversion 1.7.0
- Oozie version 5.2.0
- Phoenix version 5.0.0
- Pig version 0.17.0
- Presto version 0.238.3-amzn-1
- Presto SQL version 343
- Spark version 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow version 2.3.1
- Zeppelin version 0.9.0-preview1
- Zookeeper version 3.4.14
- Connecteurs et pilotes : Connecteur DynamoDB 4.16.0

Nouvelles fonctionnalités

- HBase: Suppression du changement de nom lors de la phase de validation et ajout d'un HFile suivi persistant. Consultez la section [HFileSuivi permanent](#) dans le guide de EMR mise à jour d'Amazon.
- HBase: Rétroporté [Créez une configuration qui oblige à mettre en cache les blocs lors du compactage](#).

- PrestoDB : améliorations apportées à l'élimination dynamique des partitions. Join Reorder basée sur des règles fonctionne sur des données non partitionnées.
- Politiques gérées délimitées : pour s'aligner sur les AWS meilleures pratiques, Amazon EMR a introduit des politiques gérées par EMR défaut définies dans la version v2 en remplacement des politiques qui seront obsolètes. Consultez les [politiques EMR gérées par Amazon](#).
- État du support du service de métadonnées d'instance (IMDS) V2 : pour Amazon EMR 6.2 ou version ultérieure, les EMR composants Amazon sont utilisés IMDSv2 pour tous les IMDS appels. Pour les IMDS appels dans le code de votre application, vous pouvez utiliser les deux IMDSv1 et IMDSv2, ou les configurer IMDS pour les utiliser uniquement IMDSv2 pour renforcer la sécurité. Si vous la désactivez IMDSv1 dans les versions antérieures d'Amazon EMR 6.x, cela entraîne un échec du démarrage du cluster.

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket

Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.

- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite du « nombre maximum de fichiers ouverts » pour les anciennes AL2 versions d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Spark : amélioration des performances dans l'environnement d'exécution de Spark.

Problèmes connus

- Amazon EMR 6.2 possède des autorisations incorrectes définies sur le fichier `libinstance-controller-java /etc/cron.d/` dans la version 6.2.0. EMR Les autorisations sur le fichier sont 645 (`-rwr--r-x`), alors qu'elles devraient être 644 (`-rw-r--r--`). Par conséquent, la EMR version 6.2 d'Amazon n'enregistre pas les journaux d'état des instances et le répertoire `/emr/instance-logs` est vide. Ce problème est résolu dans Amazon EMR 6.3.0 et versions ultérieures.

Pour contourner ce problème, exécutez le script suivant en tant qu'action d'amorçage lors du lancement du cluster.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Pour les clusters de sous-réseaux privés Amazon EMR 6.2.0 et 6.3.0, vous ne pouvez pas accéder à l'interface utilisateur Web de Ganglia. Vous recevrez un message d'erreur « accès refusé (403) ». D'autres sites WebUIs, tels que Spark, Hue JupyterHub, Zeppelin, Livy et Tez, fonctionnent normalement. L'accès à l'interface utilisateur Web de Ganglia sur les clusters de sous-réseaux publics fonctionne également normalement. Pour résoudre ce problème, redémarrez le service `httpd` sur le nœud primaire avec `sudo systemctl restart httpd`. Ce problème est résolu dans Amazon EMR 6.4.0.
- Amazon EMR 6.2.0 présente un problème où `httpd` échoue continuellement, ce qui rend Ganglia indisponible. Le message d'erreur « Impossible de se connecter au serveur » s'affiche. Pour réparer un cluster déjà en cours d'exécution présentant ce problème, SSH accédez au nœud principal du cluster et ajoutez la ligne au fichier `httpd.conf` situé `Listen 80` à l'emplacement `/etc/httpd/conf/httpd.conf`. Ce problème est résolu dans Amazon EMR 6.3.0.

- HTTPDéchoue sur les clusters EMR 6.2.0 lorsque vous utilisez une configuration de sécurité. Cela rend l'interface utilisateur de l'application web Ganglia indisponible. Pour accéder à l'interface utilisateur de l'application web Ganglia, ajoutez `Listen 80` au fichier `/etc/httpd/conf/httpd.conf` sur le nœud primaire de votre cluster. Pour plus d'informations sur la connexion à votre cluster, consultez la section [Se connecter au nœud principal à l'aide deSSH](#).

EMR Les ordinateurs portables ne parviennent pas non plus à établir de connexion avec les clusters EMR 6.2.0 lorsque vous utilisez une configuration de sécurité. Le bloc-notes ne parviendra pas à répertorier les noyaux et à soumettre les tâches Spark. Nous vous recommandons d'utiliser EMR plutôt que EMR des blocs-notes avec une autre version d'Amazon.

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` et `emr-6.2.0` sont basées sur les anciennes versions de Linux ofAmazon 2 (), qui ont un paramètre `ulimit` inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par défaut AL2. EMR AMI EMR Les versions `5.30.1`, `5.30.2`, `5.31.1`, `5.32.1`, `6.0.1`, `6.1.1`, `6.2.1`, `5.33.0`, `6.3.0` et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre `ulimit` EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre `ulimit` inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre `ulimit` lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'`ulimit` du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un `ulimit` à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section `Service`.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

Important

Amazon EMR 6.1.0 et 6.2.0 incluent un problème de performance qui peut affecter de manière critique toutes les opérations d'insertion, de modification et de suppression de Hudi. Si vous envisagez d'utiliser Hudi avec Amazon EMR 6.1.0 ou 6.2.0, vous devez contacter le AWS support pour obtenir un Hudi corrigé. RPM

• **⚠ Important**

EMR Les clusters qui exécutent Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) utilisent le comportement par défaut d'Amazon Linux et ne téléchargent ni n'installent automatiquement les mises à jour importantes et critiques du noyau nécessitant un redémarrage. Ce comportement est identique à celui des autres EC2 instances Amazon qui exécutent Amazon Linux par défaut AMI. Si de nouvelles mises à jour logicielles Amazon Linux nécessitant un redémarrage (telles que le noyau et les CUDA mises à jour) sont disponibles après la EMR sortie d'une version d'Amazon, les instances de EMR cluster qui exécutent la version par défaut AMI ne téléchargent ni n'installent automatiquement ces mises à jour. NVIDIA Pour obtenir les mises à jour du noyau, vous pouvez [personnaliser votre Amazon EMR AMI](#) pour qu'il [utilise la dernière version d'Amazon Linux AMI](#).

- Les artefacts Amazon EMR 6.2.0 Maven ne sont pas publiés. Ils seront publiés avec une future version d'AmazonEMR.
- Le HFile suivi permanent à l'aide de la table système HBase Storefile ne prend pas en charge la fonctionnalité de réplication HBase régionale. Pour plus d'informations sur HBase la réplication régionale, consultez la section Nombre [élevé de lectures disponibles cohérentes avec la chronologie](#).
- Différences entre les versions de bucket d'Amazon EMR EMR 6.x et 5.x Hive

EMR La version 5.x utilise OOS Apache Hive 2, tandis que la version EMR 6.x utilise OOS Apache Hive 3. La version open source Hive2 utilise la version 1 de Bucketing, tandis que la version open source Hive3 utilise la version 2. Cette différence de version entre Hive 2 (EMR5.x) et Hive 3 (EMR6.x) signifie que le hachage des compartiments Hive fonctionne différemment. Consultez l'exemple ci-dessous.

Le tableau suivant est un exemple créé dans les versions EMR 6.x et EMR 5.x, respectivement.

```
-- Using following LOCATION in EMR 6.x
CREATE TABLE test_bucketing (id INT, desc STRING)
PARTITIONED BY (day STRING)
CLUSTERED BY(id) INTO 128 BUCKETS
LOCATION 's3://your-own-s3-bucket/emr-6-bucketing/';

-- Using following LOCATION in EMR 5.x
LOCATION 's3://your-own-s3-bucket/emr-5-bucketing/';
```

Insérer les mêmes données dans les versions EMR 6.x et EMR 5.x.

```
INSERT INTO test_bucketing PARTITION (day='01') VALUES(66, 'some_data');
INSERT INTO test_bucketing PARTITION (day='01') VALUES(200, 'some_data');
```

La vérification de l'emplacement S3 montre que le nom du fichier de compartiment est différent, car la fonction de hachage est différente entre EMR 6.x (Hive 3) et EMR 5.x (Hive 2).

```
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-6-bucketing/day=01/
2020-10-21 20:35:16          13 000025_0
2020-10-21 20:35:22          14 000121_0
[hadoop@ip-10-0-0-122 ~]$ aws s3 ls s3://your-own-s3-bucket/emr-5-bucketing/day=01/
2020-10-21 20:32:07          13 000066_0
2020-10-21 20:32:51          14 000072_0
```

Vous pouvez également constater la différence de version en exécutant la commande suivante dans Hive CLI dans la version EMR 6.x. Notez qu'il renvoie la version 2 de compartimentation.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.

- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Version 5.31.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.31.0 d'Amazon. Les modifications ont été apportées à la version 5.30.1.

Date de parution initiale : 9 octobre 2020

Dernière mise à jour : 15 octobre 2020

Mises à niveau

- Mise à niveau du connecteur Amazon Glue vers la version 1.13.0
- Mise à niveau d'Amazon SageMaker Spark SDK vers la version 1.4.0
- Mise à niveau du connecteur Amazon Kinesis vers la version 3.5.9
- Mise à niveau AWS SDK for Java vers la version 1.11.852
- Mise à niveau de Bigtop-tomcat vers la version 8.5.56
- EMRFS mis à niveau vers la version 2.43.0
- EMR MetricsAndEventsApiGateway Client mis à niveau vers la version 1.4.0
- Mise à niveau de EMR S3 Dist CP vers la version 2.15.0
- EMRS3 Select mis à niveau vers la version 1.6.0
- Mise à niveau de Flink vers la version 1.11.0
- Mise à niveau de Hadoop vers la version 2.10.0
- Mise à niveau de Hive vers la version 2.3.7
- Mise à niveau de Hudi vers la version 0.6.0
- Mise à niveau de Hue vers la version 4.7.1
- Mise à niveau JupyterHub vers la version 1.1.0
- Mise à niveau de Mxnet vers la version 1.6.0
- Mise à niveau d'OpenCV vers la version 4.3.0
- Mise à niveau de Presto vers la version 0.238.3
- Mise à niveau TensorFlow vers la version 2.1.0

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite maximale de fichiers ouverts sur les anciennes versions AL2 d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Les [statistiques des colonnes Hive](#) sont prises en charge pour les EMR versions 5.31.0 et ultérieures d'Amazon.
- Versions de composants mises à niveau.
- EMRFS Support du S3EC V2 dans Amazon EMR 5.31.0. Dans les SDK versions 1.11.837 et ultérieures de S3 Java, le client de chiffrement version 2 (S3EC V2) a été introduit avec diverses améliorations de sécurité. Pour plus d'informations, consultez les ressources suivantes :

- Article de blog S3 : [Updates to the Amazon S3 encryption client](#).
- AWS SDK for Java Guide du développeur : [Migrez les clients de chiffrement et de déchiffrement vers la version V2](#).
- EMR Guide de gestion : [chiffrement côté client Amazon S3](#).

Le client de chiffrement V1 est toujours disponible dans le SDK pour des raisons de rétrocompatibilité.

Nouvelles fonctionnalités

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux of Amazon 2 (), qui ont un paramètre `ulimit` inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre `ulimit` EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre `ulimit` inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre `ulimit` lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'`ulimit` du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un `ulimit` à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section Service.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Avec Amazon EMR 5.31.0, vous pouvez lancer un cluster qui s'intègre à Lake Formation. Cette intégration fournit un filtrage des données précis au niveau des colonnes pour les bases de données et les tables du Glue AWS Data Catalog. Il permet également l'authentification unique fédérée à EMR Notebooks ou Apache Zeppelin à partir d'un système d'identité d'entreprise. Pour plus d'informations, consultez la section [Intégration EMR d'Amazon AWS Lake Formation](#) dans le guide EMR de gestion Amazon.

Amazon EMR with Lake Formation est actuellement disponible dans 16 AWS régions : USA Est (Ohio et Virginie du Nord), USA Ouest (Californie du Nord et Oregon), Asie-Pacifique (Mumbai, Séoul, Singapour, Sydney et Tokyo), Canada (Centre), Europe (Francfort, Irlande, Londres, Paris et Stockholm), Amérique du Sud (São Paulo).

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Lorsque AtRestEncryption HDFS le chiffrement est activé sur un cluster qui utilise Amazon EMR 5.31.0 ou 5.32.0, les requêtes Hive entraînent l'exception d'exécution suivante.

```
TaskAttempt 3 failed, info=[Error: Error while running task ( failure ) :  
attempt_1604112648850_0001_1_01_000000_3:java.lang.RuntimeException:  
java.lang.RuntimeException: Hive Runtime Error while closing  
operators: java.io.IOException: java.util.ServiceConfigurationError:
```

```
org.apache.hadoop.security.token.TokenIdentifier: Provider
org.apache.hadoop.hbase.security.token.AuthenticationTokenIdentifier not found
```

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+,-.` Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Version 6.1.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.1.0 d'Amazon. Les modifications ont été apportées à la version 6.0.0.

Date de parution initiale : 4 septembre 2020

Dernière mise à jour : 15 octobre 2020

Applications prises en charge

- AWS SDK for Java version 1.11.828
- Flink version 1.11.0
- Ganglia version 3.7.2
- Hadoop version 3.2.1-amzn-1

- HBaseversion 2.2.5
- HBase-operator-tools 1.0.0
- HCatalogversion 3.1.2-amzn-0
- Hive version 3.1.2-amzn-1
- Hudi version 0.5.2-incubation
- Hue version 4.7.1
- JupyterHub version 1.1.0
- Livy version 0.7.0
- MXNetversion 1.6.0
- Oozie version 5.2.0
- Phoenix version 5.0.0
- Presto version 0.232
- Presto SQL version 338
- Spark version 3.0.0-amzn-0
- TensorFlow version 2.1.0
- Zeppelin version 0.9.0-preview1
- Zookeeper version 3.4.14
- Connecteurs et pilotes : Connecteur DynamoDB 4.14.0

Nouvelles fonctionnalités

- ARMles types d'instance sont pris en charge à partir de la EMR version 5.30.0 d'Amazon et de la EMR version 6.1.0 d'Amazon.
- Les types d'instances à usage général m6G sont pris en charge à partir des EMR versions 6.1.0 et 5.30.0 d'Amazon. Pour plus d'informations, consultez la section [Types d'instances pris en charge](#) dans le guide EMR de gestion Amazon.
- La fonctionnalité de groupe de EC2 placement est prise en charge à partir de EMR la version 5.23.0 d'Amazon en tant qu'option pour plusieurs clusters de nœuds principaux. Actuellement, seuls les types de nœuds primaires sont pris en charge par la fonctionnalité de groupe de placement. La stratégie SPREAD est appliquée à ces nœuds primaires. La stratégie SPREAD place un petit groupe d'instances sur un matériel sous-jacent distinct afin de se prémunir contre la perte de plusieurs nœuds primaires en cas de panne matérielle. Pour plus d'informations, consultez la section [EMRIntégration au groupe de EC2 placement](#) dans le guide EMR de gestion Amazon.

- Dimensionnement géré : avec Amazon EMR version 6.1.0, vous pouvez activer le dimensionnement EMR géré par Amazon pour augmenter ou diminuer automatiquement le nombre d'instances ou d'unités dans votre cluster en fonction de la charge de travail. Amazon évalue EMR en permanence les indicateurs des clusters afin de prendre des décisions de dimensionnement qui optimisent les coûts et la rapidité de vos clusters. Managed Scaling est également disponible sur Amazon EMR version 5.30.0 et versions ultérieures, à l'exception de la version 6.0.0. Pour plus d'informations, consultez la section [Scaling Cluster Resources](#) dans le guide EMR de gestion Amazon.
- La SQL version 338 de Presto est prise en charge avec la version EMR 6.1.0. Pour plus d'informations, consultez [Presto](#).
 - Presto SQL est pris en charge sur les versions EMR 6.1.0 et ultérieures uniquement, pas sur les versions EMR 6.0.0 ou 5.x. EMR
 - Le nom de l'application, Presto continue d'être utilisé pour installer PrestoDB sur des clusters. Pour installer Presto SQL sur des clusters, utilisez le nom PrestoSQL de l'application.
 - Vous pouvez installer PrestoDB ou SQL Presto, mais vous ne pouvez pas installer les deux sur un seul cluster. Si PrestoDB et SQL Presto sont spécifiés lors de la tentative de création d'un cluster, une erreur de validation se produit et la demande de création de cluster échoue.
 - Presto SQL est pris en charge à la fois sur les clusters à maître unique et à maîtres multiples. Sur les clusters multi-maîtres, un métastore Hive externe est requis pour exécuter Presto ou PrestoDB. SQL Consultez la section [Applications prises en charge dans un EMR cluster comportant plusieurs nœuds principaux](#).
- ECR prise en charge de l'authentification automatique sur Apache Hadoop et Apache Spark avec Docker : les utilisateurs de Spark peuvent utiliser des images Docker provenant de Docker Hub et d'Amazon Elastic Container Registry (ECRAmazon) pour définir les dépendances de l'environnement et des bibliothèques.

[Configurez Docker](#) et [exécutez des applications Spark avec Docker à l'aide d'Amazon EMR 6.x](#).

- EMR prend en charge les ACID transactions Apache Hive : Amazon EMR 6.1.0 ajoute la prise en charge des ACID transactions Hive afin qu'elles soient conformes aux ACID propriétés d'une base de données. Grâce à cette fonctionnalité, vous pouvez exécuter les opérations INSERT, UPDATE, DELETE, et MERGE dans des tables gérées par Hive avec des données stockées dans Amazon Simple Storage Service (Amazon S3). Il s'agit d'une fonctionnalité clé pour les cas d'utilisation tels que l'ingestion du streaming, le retraitement des données MERGE, l'utilisation de mises à jour groupées et la modification lente des dimensions. Pour plus d'informations, notamment des

exemples de configuration et des cas d'utilisation, consultez [Amazon EMR supporte les ACID transactions Apache Hive](#).

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR
- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/ YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite maximale de fichiers ouverts sur les anciennes versions AL2 d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Apache Flink n'est pas pris en charge sur la EMR version 6.0.0, mais il l'est sur la version EMR 6.1.0 avec Flink 1.11.0. Il s'agit de la première version de Flink officiellement compatible avec Hadoop 3. Consultez [Annonce de sortie d'Apache Flink 1.11.0](#).

- Ganglia a été supprimé des ensembles de packages EMR 6.1.0 par défaut.

Problèmes connus

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux of Amazon 2 (), qui ont un paramètre ulimit inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre ulimit EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre ulimit inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre ulimit lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'ulimit du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un ulimit à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section Service.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```


Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

Important

Amazon EMR 6.1.0 et 6.2.0 incluent un problème de performance qui peut affecter de manière critique toutes les opérations d'insertion, de modification et de suppression de Hudi. Si vous envisagez d'utiliser Hudi avec Amazon EMR 6.1.0 ou 6.2.0, vous devez contacter le AWS support pour obtenir un Hudi corrigé. RPM

- Si vous définissez une configuration personnalisée de collecte des déchets avec `spark.driver.extraJavaOptions` et `spark.executor.extraJavaOptions`, cela entraînera un échec du lancement du pilote/exécuteur avec la EMR version 6.1 en raison d'une configuration de collecte des déchets conflictuelle. Avec la EMR version 6.1.0, vous devez spécifier une configuration personnalisée de collecte des déchets Spark pour les pilotes et les exécuteurs avec les propriétés `spark.driver.defaultJavaOptions` et `spark.executor.defaultJavaOptions` à la place. Pour en savoir plus, consultez

[l'environnement d'exécution Apache Spark](#) et [la configuration de Spark Garbage Collection sur Amazon EMR 6.1.0.](#)

- L'utilisation de Pig avec Oozie (et au sein de Hue, puisque Hue utilise les actions Oozie pour exécuter les scripts Pig), génère une erreur indiquant qu'une bibliothèque native-lzo ne peut pas être chargée. Ce message d'erreur est informatif et n'empêche pas Pig de fonctionner.
- Prise en charge de la simultanéité dans Hudi : Actuellement, Hudi ne prend pas en charge les écritures simultanées dans une seule table Hudi. De plus, Hudi annule toutes les modifications effectuées par les enregistreurs en cours avant de permettre à un nouvel enregistreur de démarrer. Les écritures simultanées peuvent interférer avec ce mécanisme et introduire des conditions de concurrence, ce qui peut entraîner une corruption des données. Vous devez vous assurer que, dans le cadre de votre flux de traitement des données, il n'y a qu'un seul enregistreur Hudi opérant sur une table Hudi à tout moment. Hudi prend en charge plusieurs lecteurs simultanés opérant sur la même table Hudi.
- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSH en tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Amazon EMR 6.1.0 présente un problème qui affecte les clusters exécutant Presto. Après une période prolongée (plusieurs jours), le cluster peut générer des erreurs telles que "su : failed to execute /bin/bash : Resource temporarily unavailable" ou "shell request failed on channel 0". Ce problème est dû à un EMR processus interne d'Amazon (InstanceController) qui génère trop de processus légers (LWP), ce qui finit par amener l'utilisateur Hadoop à dépasser sa limite nproc. Cela empêche l'utilisateur d'ouvrir des processus supplémentaires. La solution à ce problème consiste à effectuer une mise à niveau vers la version EMR 6.2.0.

Version 6.0.0


Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 6.0.0 d'Amazon.

Date de parution initiale : 10 mars 2020

Applications prises en charge

- AWS SDK for Java version 1.11.711
- Ganglia version 3.7.2
- Hadoop version 3.2.1
- HBaseversion 2.2.3
- HCatalogversion 3.1.2
- Hive version 3.1.2
- Hudi version 0.5.0-incubation
- Hue version 4.4.0
- JupyterHub version 1.0.0
- Livy version 0.6.0
- MXNetversion 1.5.1

- Oozie version 5.1.0
- Phoenix version 5.0.0
- Presto version 0.230
- Spark version 2.4.4
- TensorFlow version 1.14.0
- Zeppelin version 0.9.0- SNAPSHOT
- Zookeeper version 3.4.14
- Connecteurs et pilotes : Connecteur DynamoDB 4.14.0

 Note

Flink, Sqoop, Pig et Mahout ne sont pas disponibles dans la version 6.0.0 d'Amazon. EMR

Nouvelles fonctionnalités

- YARN Docker Runtime Support : YARN les applications, telles que les tâches Spark, peuvent désormais s'exécuter dans le contexte d'un conteneur Docker. Cela vous permet de définir facilement les dépendances dans une image Docker sans avoir à installer de bibliothèques personnalisées sur votre EMR cluster Amazon. Pour plus d'informations, consultez [Configurer l'intégration de Docker](#) et [exécuter des applications Spark avec Docker à l'aide d'Amazon EMR 6.0.0](#).
- LLAP Support Hive - Hive prend désormais en charge le mode LLAP d'exécution pour améliorer les performances des requêtes. Pour plus d'informations, consultez la section [Utilisation de Hive. LLAP](#)

Modifications, améliorations et problèmes résolus

- Il s'agit d'une version destinée à résoudre les problèmes liés à Amazon EMR Scaling lorsqu'il ne parvient pas à augmenter ou à réduire la taille d'un cluster ou entraîne des défaillances d'applications.
- Correction d'un problème en raison duquel les demandes de dimensionnement échouaient pour un cluster volumineux et très utilisé lorsque des EMR démons Amazon présents sur le cluster exécutaient des activités de vérification de l'état, telles que la collecte de l'état du nœud et HDFS de l'état du YARN nœud. Cela était dû au fait que les démons du cluster n'étaient pas en mesure de communiquer les données d'état de santé d'un nœud aux composants internes d'Amazon. EMR

- Démons EMR intégrés au cluster améliorés pour suivre correctement l'état des nœuds lorsque les adresses IP sont réutilisées afin d'améliorer la fiabilité lors des opérations de dimensionnement.
- [SPARK-29683](#). Correction d'un problème où les tâches échouaient lors de la réduction de la taille du cluster, car Spark supposait que tous les nœuds disponibles étaient sur la liste de refus.
- [YARN-9011](#). Correction d'un problème en raison duquel des échecs de tâches se produisaient en raison d'une course à la YARN mise hors service lorsque le cluster tentait d'augmenter ou de réduire son échelle.
- Résolution d'un problème lié aux échecs d'étapes ou de tâches lors de la mise à l'échelle du cluster en garantissant que les états des nœuds sont toujours cohérents entre les EMR démons Amazon du cluster et/. YARN HDFS
- Correction d'un problème en raison duquel les opérations de cluster telles que la réduction et la soumission par étapes échouaient pour les EMR clusters Amazon activés avec l'authentification Kerberos. Cela est dû au fait que le démon Amazon EMR on-cluster n'a pas renouvelé le ticket Kerberos, qui est nécessaire pour communiquer en toute sécurité HDFS YARN avec/exécuté sur le nœud principal.
- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite maximale de fichiers ouverts sur les anciennes versions AL2 d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Amazon Linux
 - Amazon Linux 2 est le système d'exploitation de la série de versions EMR 6.x.
 - `systemd` utilisé pour la gestion des services au lieu de `upstart` in Amazon Linux 1.
- Kit de développement Java (JDK)
 - JDK Corretto 8 est la version JDK par défaut pour la série de versions 6.x EMR.
- Scala
 - Scala 2.12 est utilisé avec Apache Spark et Apache Livy.
- Python 3
 - Python 3 est désormais la version par défaut de Python dans EMR.
- YARN Étiquettes de nœuds
 - À compter de la série de versions Amazon EMR 6.x, la fonctionnalité d'étiquettes de YARN nœuds est désactivée par défaut. Les processus principaux des applications peuvent s'exécuter à la fois sur les nœuds de noyau et sur les nœuds de tâche par défaut. Vous pouvez activer la fonctionnalité d'étiquettes de YARN nœuds en configurant les propriétés suivantes :

`yarn.node-labels.enabled` et `yarn.node-labels.am.default-node-label-expression`. Pour plus d'informations, consultez [Comprendre les nœuds primaires, les nœuds principaux et les nœuds de tâches](#).

Problèmes connus

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. Les versions Amazon : `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` et `emr-6.2.0` sont basées sur les anciennes versions de Linux of Amazon 2 (), qui ont un paramètre `ulimit` inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre `ulimit` EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. Les versions ultérieures d'Amazon AMI Le paramètre `ulimit` inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre `ulimit` lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'`ulimit` du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un `ulimit` à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section `Service`.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer `InstanceController`

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Le shell interactif Spark PySpark, y compris SparkR et spark-shell, ne prend pas en charge l'utilisation de Docker avec des bibliothèques supplémentaires.
- Pour utiliser Python 3 avec Amazon EMR version 6.0.0, vous devez ajouter `PATH` à `yarn.nodemanager.env-whitelist`.
- Les fonctionnalités Live Long et Process (LLAP) ne sont pas prises en charge lorsque vous utilisez le catalogue de données AWS Glue comme métastore pour Hive.
- Lorsque vous utilisez Amazon EMR 6.0.0 avec l'intégration de Spark et Docker, vous devez configurer les instances de votre cluster avec le même type d'instance et le même nombre de EBS volumes afin d'éviter toute défaillance lors de la soumission d'une tâche Spark avec Docker Runtime.
- Dans Amazon EMR 6.0.0, HBase le mode de stockage Amazon S3 est concerné par le problème [HBASE-24286](#). HBasemaster ne peut pas s'initialiser lorsque le cluster est créé à l'aide de données S3 existantes.

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.30.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.30.1 d'Amazon. Les modifications ont été apportées à la version 5.30.0.

Date de parution initiale : 30 juin 2020

Dernière mise à jour : 24 août 2020

Modifications, améliorations et problèmes résolus

- Les nouvelles EMR versions d'Amazon résolvent le problème en abaissant la limite maximale de fichiers ouverts sur les anciennes versions AL2 d'AmazonEMR. EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent désormais un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé.
- Correction d'un problème à cause duquel le processus du contrôleur d'instance générait un nombre infini de processus.
- Correction d'un problème où Hue ne parvenait pas à exécuter une requête Hive, affichant un message « la base de données est verrouillée » et empêchant l'exécution des requêtes.
- Correction d'un problème lié à Spark qui permettait à un plus grand nombre de tâches de s'exécuter simultanément sur le EMR cluster.
- Correction d'un problème de bloc-notes Jupyter provoquant une erreur « trop de fichiers ouverts » dans le serveur Jupyter.
- Correction d'un problème lié aux temps de démarrage des clusters.

Nouvelles fonctionnalités

- Les interfaces d'application persistantes de Tez UI et de YARN Timeline Server sont disponibles avec EMR les versions 6.x, 5.30.1 et EMR ultérieures d'Amazon. L'accès par lien en un clic à l'historique permanent des candidatures vous permet d'accéder rapidement à l'historique des tâches sans configurer de proxy Web via une SSH connexion. Les journaux des clusters actifs et résiliés sont disponibles pendant 30 jours après la fin de l'application. Pour plus d'informations, consultez la section [Afficher les interfaces utilisateur persistantes des applications](#) dans le Amazon EMR Management Guide.
- EMR L'exécution APIs de blocs-notes est disponible pour exécuter EMR des blocs-notes via un script ou une ligne de commande. La possibilité de démarrer, d'arrêter, de répertorier et de décrire les exécutions de EMR blocs-notes sans la AWS console vous permet de contrôler un EMR bloc-notes par programme. À l'aide d'une cellule de bloc-notes paramétrée, vous pouvez transmettre différentes valeurs de paramètres à un bloc-notes sans avoir à créer une copie du bloc-notes pour chaque nouvel ensemble de valeurs de paramètres. Voir [EMRAPIActions](#). Pour un exemple de code, consultez la section [Exemples de commandes pour exécuter des EMR blocs-notes par programmation](#).

Problèmes connus

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux of Amazon 2 (), qui ont un paramètre ulimit inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre ulimit EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre ulimit inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre ulimit lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'ulimit du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un ulimit à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section Service.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- EMRCarnets

La fonctionnalité qui permet d'installer des noyaux et des bibliothèques Python supplémentaires sur le nœud principal du cluster est désactivée par défaut dans la EMR version 5.30.1. Pour plus d'informations sur cette fonctionnalité, consultez [Installation de noyaux et de bibliothèques Python sur un nœud primaire de cluster](#).

Pour activer cette fonctionnalité, procédez comme suit :

1. Assurez-vous que la politique d'autorisation associée au rôle de service pour les EMR blocs-notes autorise les actions suivantes :

```
elasticmapreduce:ListSteps
```

Pour plus d'informations, consultez la section [Rôle de service pour les EMR ordinateurs portables](#).

2. Utilisez le AWS CLI pour exécuter une étape sur le cluster qui configure les EMR blocs-notes, comme indiqué dans l'exemple suivant. Remplacez *us-east-1* avec la région dans laquelle

réside votre cluster. Pour plus d'informations sur l'ajout d'étapes, consultez la rubrique [Ajout d'étapes à un cluster à l'aide de la AWS CLI](#).

```
aws emr add-steps --cluster-id MyClusterID --steps
  Type=CUSTOM_JAR,Name=EMRNotebooksSetup,ActionOnFailure=CONTINUE,Jar=s3://us-
  east-1.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
  awssupportdatasvcs.com/bootstrap-actions/EMRNotebooksSetup/emr-notebooks-setup.sh"]
```

- Mise à l'échelle gérée

Les opérations de mise à l'échelle gérées sur des clusters 5.30.0 et 5.30.1 sans Presto installé peuvent provoquer des défaillances d'applications ou empêcher le maintien d'un groupe d'instances ou d'une flotte d'instances uniforme dans l'état ARRESTED, en particulier lorsqu'une opération de réduction est rapidement suivie d'une opération d'augmentation.

Pour contourner le problème, choisissez Presto comme application à installer lorsque vous créez un cluster avec les EMR versions 5.30.0 et 5.30.1 d'Amazon, même si votre travail ne nécessite pas Presto.

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
- Deux partitions ou plus sont analysées à partir de la même table.
- Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
- Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre a et b dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez qu'il existe 14 autres caractères de non-contrôle : `!"#$%&'()*+, -`. Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Version 5.30.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.30.0 d'Amazon. Les modifications ont été apportées à la version 5.29.0.

Date de parution initiale : 13 mai 2020

Date de la dernière mise à jour : 25 juin 2020

Mises à niveau

- Mise à niveau AWS SDK for Java vers la version 1.11.759
- Mise à niveau d'Amazon SageMaker Spark SDK vers la version 1.3.0
- Serveur d'EMR enregistrement mis à niveau vers la version 1.6.0
- Mise à niveau de Flink vers la version 1.10.0
- Mise à niveau de Ganglia vers la version 3.7.2
- Mise à niveau HBase vers la version 1.4.13
- Mise à niveau de Hudi vers la version 0.5.2 incubating
- Mise à niveau de Hue vers la version 4.6.0
- Mise à niveau JupyterHub vers la version 1.1.0
- Mise à niveau de Livy vers la version 0.7.0-incubating
- Mise à niveau d'Oozie vers la version 5.2.0
- Mise à niveau de Presto vers la version 0.232
- Mise à niveau de Spark vers la version 2.4.5
- Connecteurs et pilotes améliorés : Amazon Glue Connector 1.12.0 ; Amazon Kinesis Connector 3.5.0 ; DynamoDB Connector 4.14.0 EMR

Nouvelles fonctionnalités

- EMR Blocs-notes — Lorsqu'ils sont utilisés avec des EMR clusters créés à l'aide de la version 5.30.0, les noyaux des EMR blocs-notes s'exécutent sur le cluster. Cela améliore les performances des blocs-notes et vous permet d'installer et de personnaliser les noyaux. Vous pouvez également installer des bibliothèques Python sur le nœud primaire du cluster. Pour plus d'informations, consultez la section [Installation et utilisation des noyaux et des bibliothèques](#) dans le Guide EMR de gestion.
- Dimensionnement géré : avec Amazon EMR version 5.30.0 et versions ultérieures, vous pouvez activer le dimensionnement EMR géré pour augmenter ou diminuer automatiquement le nombre d'instances ou d'unités dans votre cluster en fonction de la charge de travail. Amazon évalue EMR en permanence les indicateurs des clusters afin de prendre des décisions de dimensionnement qui optimisent les coûts et la rapidité de vos clusters. Pour plus d'informations, consultez la section [Scaling Cluster Resources](#) dans le guide EMR de gestion Amazon.
- Chiffrer les fichiers journaux stockés dans Amazon S3 : avec Amazon EMR version 5.30.0 et versions ultérieures, vous pouvez chiffrer les fichiers journaux stockés dans Amazon S3 à l'aide

d'une AWS KMS clé gérée par le client. Pour plus d'informations, consultez la section [Chiffrer les fichiers journaux stockés dans Amazon S3](#) dans le guide EMR de gestion Amazon.

- Support d'Amazon Linux 2 : dans les EMR versions 5.30.0 et ultérieures, EMR uses Amazon système d'exploitation Linux 2. La nouvelle personnalisation AMIs (Amazon Machine Image) doit être basée sur the Amazon Linux 2 AMI. Pour plus d'informations, consultez la section [Utilisation d'une option personnalisée AMI](#).
- Presto Graceful Auto Scale : les EMR clusters utilisant la version 5.30.0 peuvent être définis avec un délai de mise à l'échelle automatique qui donne aux tâches Presto le temps de terminer leur exécution avant que leur nœud ne soit mis hors service. Pour de plus amples informations, veuillez consulter [Utilisation du dimensionnement automatique de Presto avec désaffectation gracieuse](#).
- Création d'une instance de flotte avec une nouvelle option de stratégie d'allocation — Une nouvelle option de stratégie d'allocation est disponible dans les EMR versions 5.12.1 et ultérieures. Cela permet un approvisionnement plus rapide des clusters, une allocation plus précise et moins d'interruptions d'instances Spot. Des mises à jour des rôles de EMR service autres que ceux par défaut sont requises. Consultez [Configuration de parcs d'instances](#).
- Commandes `sudo systemctl stop` et `sudo systemctl start` : dans les EMR versions 5.30.0 et ultérieures, quel système d'exploitation use Amazon Linux 2 utilise et commande pour redémarrer les services. `EMR sudo systemctl stop sudo systemctl start` Pour plus d'informations, consultez [Comment redémarrer un service sur Amazon EMR ?](#).

Modifications, améliorations et problèmes résolus

- EMR la version 5.30.0 n'installe pas Ganglia par défaut. Lorsque vous créez un cluster, vous pouvez sélectionner expressément l'installation de Ganglia.
- Optimisation des performances Spark.
- Optimisation des performances Presto.
- Python 3 est la EMR version par défaut d'Amazon 5.30.0 et versions ultérieures.
- Le groupe de sécurité géré par défaut pour l'accès au service dans les sous-réseaux privés a été mis à jour avec de nouvelles règles. Si vous utilisez un groupe de sécurité personnalisé pour accéder au service, vous devez inclure les mêmes règles que le groupe de sécurité géré par défaut. Pour plus d'informations, consultez [Amazon EMR -Managed Security Group for Service Access \(sous-réseaux privés\)](#). Si vous utilisez un rôle de service personnalisé pour Amazon EMR, vous devez accorder l'autorisation `ec2:describeSecurityGroups` EMR afin de valider si les groupes de sécurité ont été correctement créés. Si vous utilisez le `EMR_DefaultRole`, cette autorisation est déjà incluse dans la stratégie gérée par défaut.

Problèmes connus

- Réduction de la limite du « nombre maximum de fichiers ouverts » pour les anciennes versions AL2 [corrigée dans les nouvelles versions]. EMR Les versions Amazon : emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 et emr-6.2.0 sont basées sur les anciennes versions de Linux of Amazon 2 (), qui ont un paramètre ulimit inférieur pour le « Nombre maximum de fichiers ouverts » lorsque les clusters Amazon sont créés avec la valeur par AL2 défaut. EMR AMI EMR Les versions 5.30.1, 5.30.2, 5.31.1, 5.32.1, 6.0.1, 6.1.1, 6.2.1, 5.33.0, 6.3.0 et versions ultérieures d'Amazon incluent un correctif permanent avec un paramètre « Nombre maximum de fichiers ouverts » plus élevé. Les versions dont la limite de fichiers ouverts est inférieure provoquent l'erreur « Trop de fichiers ouverts » lors de la soumission d'une tâche Spark. Dans les versions concernées, le paramètre ulimit EMR par défaut d'Amazon AMI est de 4096 pour le « Nombre maximum de fichiers ouverts », ce qui est inférieur à la limite de fichiers de 65536 sous Linux 2. latest Amazon AMI Le paramètre ulimit inférieur pour « Nombre maximum de fichiers ouverts » entraîne l'échec de la tâche Spark lorsque le pilote et l'exécuteur Spark tentent d'ouvrir plus de 4 096 fichiers. Pour résoudre ce problème, Amazon EMR dispose d'un script d'action bootstrap (BA) qui ajuste le paramètre ulimit lors de la création du cluster.

Si vous utilisez une ancienne EMR version d'Amazon qui ne contient pas de solution permanente à ce problème, la solution suivante vous permet de définir explicitement l'ulimit du contrôleur d'instance sur un maximum de 65536 fichiers.

Définir explicitement un ulimit à partir de la ligne de commande

1. Modifiez `/etc/systemd/system/instance-controller.service` pour ajouter les paramètres suivants à la section Service.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Redémarrer InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Définissez un ulimit à l'aide de l'action d'amorçage (BA)

Vous pouvez également utiliser un script d'action d'amorçage (BA) pour configurer ulimit du contrôleur d'instance à 65536 fichiers lors de la création du cluster.

```
#!/bin/bash
for user in hadoop spark hive; do
sudo tee /etc/security/limits.d/$user.conf << EOF
$user - nofile 65536
$user - nproc 65536
EOF
done
for proc in instancecontroller logpusher; do
sudo mkdir -p /etc/systemd/system/$proc.service.d/
sudo tee /etc/systemd/system/$proc.service.d/override.conf << EOF
[Service]
LimitNOFILE=65536
LimitNPROC=65536
EOF
pid=$(pgrep -f aws157.$proc.Main)
sudo prlimit --pid $pid --nofile=65535:65535 --nproc=65535:65535
done
sudo systemctl daemon-reload
```

- Mise à l'échelle gérée

Les opérations de mise à l'échelle gérées sur des clusters 5.30.0 et 5.30.1 sans Presto installé peuvent provoquer des défaillances d'applications ou empêcher le maintien d'un groupe d'instances ou d'une flotte d'instances uniforme dans l'état ARRESTED, en particulier lorsqu'une opération de réduction est rapidement suivie d'une opération d'augmentation.

Pour contourner le problème, choisissez Presto comme application à installer lorsque vous créez un cluster avec les EMR versions 5.30.0 et 5.30.1 d'Amazon, même si votre travail ne nécessite pas Presto.

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois

sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

- Le moteur de base de données par défaut pour Hue 4.6.0 est SQLite, ce qui pose des problèmes lorsque vous essayez d'utiliser Hue avec une base de données externe. Pour résoudre ce problème, définissez `engine` de votre configuration de classification `hue.ini` sur `mysql`. Ce problème a été résolu dans la EMR version 5.30.1 d'Amazon.
- Lorsque vous utilisez Spark avec le formatage de l'emplacement des partitions Hive pour lire des données dans Amazon S3 et que vous exécutez Spark sur Amazon EMR versions 5.30.0 à 5.36.0 et 6.2.0 à 6.9.0, vous pouvez rencontrer un problème empêchant votre cluster de lire correctement les données. Cela peut se produire si vos partitions présentent toutes les caractéristiques suivantes :
 - Deux partitions ou plus sont analysées à partir de la même table.
 - Au moins un chemin de répertoire de partition est un préfixe d'au moins un autre chemin de répertoire de partition, par exemple, `s3://bucket/table/p=a` est un préfixe de `s3://bucket/table/p=a b`.
 - Le premier caractère qui suit le préfixe dans l'autre répertoire de partitions possède une valeur UTF -8 inférieure au / caractère (U+002F). Par exemple, le caractère d'espace (U+0020) qui apparaît entre `a` et `b` dans `s3://bucket/table/p=a b` entre dans cette catégorie. Notez

qu'il existe 14 autres caractères de non-contrôle : !"#%&'()*+,-. Pour plus d'informations, consultez la [table de codage UTF -8 et les caractères Unicode](#).

Pour contourner ce problème, définissez la configuration `spark.sql.sources.fastS3PartitionDiscovery.enabled` sur `false` dans la classification `spark-defaults`.

Version 5.29.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.29.0 d'Amazon. Les modifications ont été apportées à la version 5.28.1.

Date de parution initiale : 17 janvier 2020

Mises à niveau

- Mise à niveau AWS SDK for Java vers la version 1.11.682
- Mise à niveau de Hive vers la version 2.3.6
- Mise à niveau de Flink vers la version 1.9.1
- Mise à niveau d'EmrFS vers la version 2.38.0
- Connecteur EMR DynamoDB mis à niveau vers la version 4.13.0

Modifications, améliorations et problèmes résolus

- Spark
 - Optimisation des performances Spark.
- EMRFS
 - Mise à jour du guide de gestion avec les paramètres par défaut du fichier `emrfs-site.xml` pour un affichage cohérent.

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes

avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.28.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.28.1 d'Amazon. Les modifications ont été apportées à la version 5.28.0.

Date de parution initiale : 10 janvier 2020

Modifications, améliorations et problèmes résolus

- Spark
 - Correction des problèmes de compatibilité Spark.
- CloudWatch Métriques
 - Correction de la publication d'Amazon CloudWatch Metrics sur un EMR cluster comportant plusieurs nœuds principaux.

- Message de journal désactivé
 - Message de faux journal désactivé, « ...en utilisant l'ancienne version (<4.5.8) du client HTTP Apache ».

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.28.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.28.0 d'Amazon. Les modifications ont été apportées à la version 5.27.0.

Date de parution initiale : 12 novembre 2019

Mises à niveau

- Mise à niveau de Flink vers la version 1.9.0
- Mise à niveau de Hive vers la version 2.3.6
- Mise à niveau MXNet vers la version 1.5.1
- Mise à niveau de Phoenix vers la version 4.14.3
- Mise à niveau de Presto vers la version 0.227
- Mise à niveau de Zeppelin vers la version 0.8.2

Nouvelles fonctionnalités

- Amazon EMR peut désormais installer [Apache Hudi](#) lorsque vous créez un cluster. Pour de plus amples informations, veuillez consulter [Hudi](#).
- (25 novembre 2019) Vous pouvez maintenant choisir d'exécuter plusieurs étapes en parallèle pour améliorer l'utilisation du cluster et faire des économies. Vous pouvez également annuler à la fois les étapes en attente et celles en cours d'exécution. Pour plus d'informations, voir [Travailler avec des étapes à l'aide de la console AWS CLI AND](#).
- (3 décembre 2019) Vous pouvez désormais créer et exécuter des EMR clusters sur AWS Outposts. AWS Outposts permet AWS des services, une infrastructure et des modèles d'exploitation natifs dans les installations sur site. Dans AWS Outposts les environnements, vous pouvez utiliser les mêmes AWS APIs outils et infrastructures que ceux que vous utilisez dans le AWS cloud. Pour plus d'informations, consultez la section [EMRclusters sur AWS Outposts](#).
- (11 mars 2020) À partir de la EMR version 5.28.0 d'Amazon, vous pouvez créer et exécuter des EMR clusters Amazon sur un sous-réseau de zones AWS locales en tant qu'extension logique d'une AWS région qui prend en charge les zones locales. Une zone locale permet aux EMR fonctionnalités d'Amazon et à un sous-ensemble de AWS services, tels que les services de calcul et de stockage, d'être situés plus près des utilisateurs, offrant ainsi un accès à très faible latence aux applications exécutées localement. Pour obtenir la liste des zones locales

disponibles, consultez [AWS Local Zones](#). Pour plus d'informations sur l'accès aux zones AWS locales disponibles, voir [Régions, zones de disponibilité et zones locales](#).

Les zones Locales ne prennent actuellement pas en charge les EMR ordinateurs portables Amazon et ne prennent pas en charge les connexions directes à Amazon EMR via l'interface VPC endpoint (AWS PrivateLink).

Modifications, améliorations et problèmes résolus

- Prise en charge d'applications étendues pour des clusters à haute disponibilité
 - Pour plus d'informations, consultez la section [Applications prises en charge dans un EMR cluster comportant plusieurs nœuds principaux](#) dans le guide EMR de gestion Amazon.
- Spark
 - Optimisation des performances
- Hive
 - Optimisation des performances
- Presto
 - Optimisation des performances

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.27.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.27.0 d'Amazon. Les modifications ont été apportées à la version 5.26.0.

Date de parution initiale : 23 septembre 2019

Mises à niveau

- AWS SDK for Java 1,1,615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- TensorFlow 1.14.0
- Connecteurs et pilotes :
 - Connecteur DynamoDB 4.12.0

Nouvelles fonctionnalités

- (24 octobre 2019) Les nouvelles fonctionnalités suivantes des EMR ordinateurs portables sont disponibles avec toutes les EMR versions d'Amazon.
 - Vous pouvez désormais associer des référentiels Git à des EMR blocs-notes pour stocker vos blocs-notes dans un environnement à version contrôlée. Vous pouvez partager du code avec

vos pairs et réutiliser les blocs-notes Jupyter existants via des référentiels Git distants. Pour plus d'informations, consultez [Associer des référentiels Git à Amazon EMR Notebooks](#) dans le guide de EMRgestion Amazon.

- L'[utilitaire nbdime](#) est désormais disponible dans les ordinateurs portables pour simplifier la comparaison et la fusion EMR des blocs-notes.
- EMRles ordinateurs portables sont désormais compatibles JupyterLab. JupyterLab est un environnement de développement interactif basé sur le Web entièrement compatible avec les ordinateurs portables Jupyter. Vous pouvez désormais choisir d'ouvrir votre bloc-notes dans l'éditeur de bloc-notes Jupyter JupyterLab ou dans l'éditeur de bloc-notes Jupyter.
- (30 octobre 2019) Avec EMR les versions 5.25.0 et ultérieures d'Amazon, vous pouvez vous connecter à l'interface utilisateur du serveur d'historique Spark depuis la page de résumé du cluster ou depuis l'onglet Historique des applications de la console. Au lieu de configurer un proxy Web via une SSH connexion, vous pouvez accéder rapidement à l'interface utilisateur du serveur d'historique Spark pour consulter les métriques de l'application et accéder aux fichiers journaux pertinents pour les clusters actifs et fermés. Pour plus d'informations, consultez la section [Accès hors cluster aux interfaces utilisateur persistantes des applications](#) dans le Amazon EMR Management Guide.

Modifications, améliorations et problèmes résolus

- EMRCluster Amazon avec plusieurs nœuds principaux
 - Vous pouvez installer et exécuter Flink sur un EMR cluster Amazon comportant plusieurs nœuds principaux. Pour plus d'informations, consultez [Applications et fonctionnalités prises en charge](#).
 - Vous pouvez configurer le chiffrement HDFS transparent sur un EMR cluster Amazon comportant plusieurs nœuds principaux. Pour plus d'informations, consultez la section [Chiffrement HDFS transparent sur les EMR clusters comportant plusieurs nœuds principaux](#).
 - Vous pouvez désormais modifier la configuration des applications exécutées sur un EMR cluster Amazon comportant plusieurs nœuds principaux. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).
- Connecteur Amazon EMR -DynamoDB
 - Le connecteur Amazon EMR -DynamoDB prend désormais en charge les types de données DynamoDB suivants : boolean, list, map, item, null. Pour plus d'informations, consultez [Configuration d'une table Hive pour l'exécution de commandes Hive](#).

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.26.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.26.0 d'Amazon. Les modifications ont été apportées à la version 5.25.0.

Date de parution initiale : 8 août 2019

Dernière mise à jour : 19 août 2019

Mises à niveau

- AWS SDK for Java 1,1595
- HBase1.4.10
- Phoenix 4.14.2
- Connecteurs et pilotes :
 - Connecteur DynamoDB 4.11.0
 - Connecteur MariaDB 2.4.2
 - Pilote Amazon Redshift JDBC 1.2.32.1056

Nouvelles fonctionnalités

- (Bêta) Avec Amazon EMR 5.26.0, vous pouvez lancer un cluster qui s'intègre à Lake Formation. Cette intégration fournit un accès détaillé au niveau des colonnes aux bases de données et aux tables du Glue AWS Data Catalog. Il permet également l'authentification unique fédérée à EMR Notebooks ou Apache Zeppelin à partir d'un système d'identité d'entreprise. Pour plus d'informations, consultez [Intégrer Amazon EMR à AWS Lake Formation \(version bêta\)](#).
- (19 août 2019) L'accès public EMR bloqué par Amazon est désormais disponible dans toutes les EMR versions d'Amazon qui prennent en charge les groupes de sécurité. Bloquer l'accès public est un paramètre applicable à l'ensemble du compte et appliqué à chaque AWS région. Le blocage de l'accès public empêche le lancement d'un cluster lorsqu'un groupe de sécurité associé au cluster possède une règle autorisant le trafic entrant depuis IPv4 0.0.0.0/0 ou IPv6 :::/0 (accès public) sur un port, sauf si un port est spécifié comme exception. Le port 22 est une exception par défaut. Pour plus d'informations, consultez la section [Utilisation d'Amazon EMR Block Public Access](#) dans le guide EMR de gestion Amazon.

Modifications, améliorations et problèmes résolus

- EMRCarnets
 - À partir de la version EMR 5.26.0, EMR Notebooks prend en charge les bibliothèques Python adaptées aux ordinateurs portables en plus des bibliothèques Python par défaut. Vous pouvez installer des bibliothèques limitées au bloc-notes depuis l'éditeur de bloc-notes sans avoir à recréer un cluster ou à rattacher un bloc-notes à un cluster. Les bibliothèques limitées aux blocs-notes sont créées dans un environnement virtuel Python, de sorte qu'elles ne s'appliquent qu'à la

session de bloc-notes en cours. Cela vous permet d'isoler les dépendances du bloc-notes. Pour plus d'informations, consultez la section [Using Notebook Scoped Libraries](#) dans le Amazon EMR Management Guide.

- EMRFS

- Vous pouvez activer une fonctionnalité ETag de vérification (version bêta) en réglant `fs.s3.consistent.metadata.etag.validation.enabled` sur `true`. Grâce à cette fonctionnalité, EMRFS utilise Amazon S3 ETags pour vérifier que les objets en cours de lecture sont de la dernière version disponible. Cette fonctionnalité est utile dans les cas read-after-update d'utilisation dans lesquels des fichiers sur Amazon S3 sont remplacés tout en conservant le même nom. Cette fonctionnalité de ETag vérification ne fonctionne pas actuellement avec S3 Select. Pour plus d'informations, consultez [Configuration de la vue cohérente](#).

- Spark

- Les optimisations suivantes sont désormais activées par défaut : élagage dynamique des partitions, amélioration de l'inférence des statistiques DISTINCT du SQL plan pour les requêtes JOIN suivies INTERSECT, aplatissement des DISTINCT sous-requêtes scalaires, réorganisation optimisée des jointures et jointure par filtre Bloom. Pour plus d'informations, consultez [Optimisation des performances de Spark](#).
- Génération de code améliorée pour l'ensemble de l'étape pour Sort Merge Join.
- Amélioration de la réutilisation des fragments de requête et des sous-requêtes.
- Améliorations apportées à la préallocation des exécuteurs au démarrage de Spark.
- Les jointures par filtre bloom ne sont plus appliquées lorsque le côté le plus petit de la jointure inclut un indice de diffusion.

- Tez

- Un problème avec Tez a été résolu. L'interface utilisateur de Tez fonctionne désormais sur un EMR cluster Amazon avec plusieurs nœuds principaux.

Problèmes connus

- Les fonctionnalités améliorées de génération de code pour l'ensemble de l'étape pour Sort Merge Join peuvent augmenter la sollicitation de la mémoire lorsqu'elles sont activées. Cette optimisation améliore les performances, mais peut entraîner de nouvelles tentatives de tâches ou des échecs si `spark.yarn.executor.memoryOverheadFactor` n'est pas réglé pour fournir suffisamment de mémoire. Pour désactiver cette fonctionnalité, définissez `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` sur `false`.

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.25.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.25.0 d'Amazon. Les modifications ont été apportées à la version 5.24.1.

Date de parution initiale : 17 juillet 2019

Dernière mise à jour : 30 octobre 2019

Amazon EMR 5.25.0

Mises à niveau

- AWS SDK for Java 1,1,566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3
- TensorFlow 1.13.1
- Tez 0.9.2
- Zookeeper 3.4.14

Nouvelles fonctionnalités

- (30 octobre 2019) À partir de la EMR version 5.25.0 d'Amazon, vous pouvez vous connecter à l'interface utilisateur du serveur d'historique Spark depuis la page de résumé du cluster ou depuis l'onglet Historique des applications de la console. Au lieu de configurer un proxy Web via une SSH connexion, vous pouvez accéder rapidement à l'interface utilisateur du serveur d'historique Spark pour consulter les métriques de l'application et accéder aux fichiers journaux pertinents pour les clusters actifs et fermés. Pour plus d'informations, consultez la section [Accès hors cluster aux interfaces utilisateur persistantes des applications](#) dans le Amazon EMR Management Guide.

Modifications, améliorations et problèmes résolus

- Spark
 - Amélioration des performances de certaines jointures en utilisant des filtres Bloom pour préfiltrer les entrées. L'optimisation est désactivée par défaut et peut être activée en définissant le paramètre de configuration Spark `spark.sql.bloomFilterJoin.enabled` sur `true`.
 - Amélioration des performances du regroupement par colonnes de type chaîne.
 - Amélioration de la mémoire par défaut de l'exécuteur Spark et de la configuration des cœurs des types d'instances R4 pour les clusters non HBase installés.
 - Résolution d'un problème antérieur lié à la fonctionnalité d'élimination dynamique des partitions, où la table éliminée devait se trouver du côté gauche de la jointure.
 - Amélioré DISTINCT avant INTERSECT l'optimisation pour s'appliquer à d'autres cas impliquant des alias.

- Amélioration de l'inférence des statistiques du SQL plan pour les requêtes JOIN suivies par les DISTINCT requêtes. Cette amélioration est désactivée par défaut et peut être activée en définissant le paramètre de configuration Spark `spark.sql.statsImprovements.enabled` sur `true`. Cette optimisation est requise par la fonctionnalité Distinct avant Intersect et sera automatiquement activée lorsque `spark.sql.optimizer.distinctBeforeIntersect.enabled` est définie sur `true`.
- Ordre de jointure optimisé en fonction de la taille de la table et des filtres. Cette optimisation est désactivée par défaut et peut être activée en définissant le paramètre de configuration Spark `spark.sql.optimizer.sizeBasedJoinReorder.enabled` sur `true`.

Pour plus d'informations, consultez [Optimisation des performances de Spark](#).

- EMRFS
 - Le EMRFS paramètre est désormais désactivé par défaut. `fs.s3.buckets.create.enabled` Lors des tests, nous avons constaté que la désactivation de ce paramètre améliore les performances et empêche la création involontaire de compartiments S3. Si votre application repose sur cette fonctionnalité, vous pouvez l'activer en définissant la propriété `fs.s3.buckets.create.enabled` sur `true` dans la classification de configuration `emrfs-site`. Pour plus d'informations, consultez [Fourniture d'une configuration lors de la création d'un cluster](#).
- Améliorations du chiffrement du disque local et du chiffrement S3 dans les configurations de sécurité (5 août 2019)
 - Les paramètres de chiffrement Amazon S3 ont été séparés des paramètres de chiffrement du disque local dans la configuration de la sécurité.
 - Ajout d'une option pour activer EBS le chiffrement avec les versions 5.24.0 et ultérieures. La sélection de cette option chiffre le volume du périphérique racine en plus des volumes de stockage. Les versions précédentes nécessitaient l'utilisation d'une AMI méthode personnalisée pour chiffrer le volume du périphérique racine.
 - Pour plus d'informations, consultez la section [Options de chiffrement](#) dans le guide EMR de gestion Amazon.

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.24.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.24.1 d'Amazon. Les modifications ont été apportées à la version 5.24.0.

Date de parution initiale : 26 juin 2019

Modifications, améliorations et problèmes résolus

- Mise à jour de la version par défaut d'Amazon Linux AMI pour Amazon EMR afin d'inclure d'importantes mises à jour de sécurité du noyau Linux, notamment le problème de TCP SACK déni de service ([AWS-2019-005](#)).

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.24.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.24.0 d'Amazon. Les modifications ont été apportées à la version 5.23.0.

Date de parution initiale : 11 juin 2019

Dernière mise à jour : 5 août 2019

Mises à niveau

- Flink 1.8.0
- Hue 4.4.0
- JupyterHub 0,9,6
- Livy 0.6.0
- MxNet 1.4.0
- Presto 0.219
- Spark 2.4.2
- AWS SDK for Java 1,1,546
- Connecteurs et pilotes :
 - Connecteur DynamoDB 4.9.0
 - Connecteur MariaDB 2.4.1
 - Pilote Amazon Redshift JDBC 1.2.27.1051

Modifications, améliorations et problèmes résolus

- Spark
 - Ajout d'une optimisation pour éliminer dynamiquement les partitions. L'optimisation est désactivée par défaut. Pour l'activer, définissez le paramètre de configuration Spark `spark.sql.dynamicPartitionPruning.enabled` sur `true`.
 - Amélioration des performances des requêtes INTERSECT. L'optimisation est désactivée par défaut. Pour l'activer, définissez le paramètre de configuration Spark `spark.sql.optimizer.distinctBeforeIntersect.enabled` sur `true`.
 - Ajout d'une optimisation pour aplatir les sous-requêtes scalaires avec des agrégats qui utilisent la même relation. L'optimisation est désactivée

par défaut. Pour l'activer, définissez le paramètre de configuration Spark `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` sur `true`.

- Génération de code améliorée pour l'ensemble de l'étape.

Pour plus d'informations, consultez [Optimisation des performances de Spark](#).

- Améliorations du chiffrement du disque local et du chiffrement S3 dans les configurations de sécurité (5 août 2019)
 - Les paramètres de chiffrement Amazon S3 ont été séparés des paramètres de chiffrement du disque local dans la configuration de la sécurité.
 - Ajout d'une option pour activer EBS le chiffrement. La sélection de cette option chiffre le volume du périphérique racine en plus des volumes de stockage. Les versions précédentes nécessitaient l'utilisation d'une AMI méthode personnalisée pour chiffrer le volume du périphérique racine.
 - Pour plus d'informations, consultez la section [Options de chiffrement](#) dans le guide EMR de gestion Amazon.

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.23.0 d'Amazon. Les modifications ont été apportées à la version 5.22.0.

Date de parution initiale : 1 avril 2019

Dernière mise à jour : 30 avril 2019

Mises à niveau

- AWS SDK for Java 1,1,519

Nouvelles fonctionnalités

- (30 avril 2019) Avec Amazon EMR 5.23.0 et versions ultérieures, vous pouvez lancer un cluster avec trois nœuds principaux pour prendre en charge la haute disponibilité d'applications telles que YARN Resource Manager, Spark HDFS NameNode, Hive et Ganglia. Le nœud primaire n'est plus un point de défaillance potentiel grâce à cette fonctionnalité. Si l'un des nœuds principaux tombe en panne, Amazon EMR bascule automatiquement vers un nœud principal de secours et remplace le nœud principal défaillant par un nouveau nœud présentant les mêmes actions de configuration et de démarrage. Pour plus d'informations, consultez [Planification et configuration des nœuds primaires](#).

Problèmes connus

- Interface utilisateur Tez (corrigée dans la EMR version 5.26.0 d'Amazon)

L'interface utilisateur de Tez ne fonctionne pas sur un EMR cluster comportant plusieurs nœuds principaux.

- Hue (corrigé dans la EMR version 5.24.0 d'Amazon)
 - Hue exécuté sur Amazon EMR ne prend pas en charge Solr. À compter de la EMR version 5.20.0 d'Amazon, un problème de configuration incorrecte entraîne l'activation de Solr et un message d'erreur inoffensif similaire au suivant s'affiche :

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

Pour empêcher l'affichage du message d'erreur de Solr, procédez comme suit :

1. Connectez-vous à la ligne de commande du nœud principal à l'aide deSSH.
2. Utilisez un éditeur de texte pour ouvrir le fichier `hue.ini`. Par exemple :

```
sudo vim /etc/hue/conf/hue.ini
```

3. Recherchez le terme `appblacklist` et modifiez la ligne comme suit :

```
appblacklist = search
```

4. Enregistrez vos modifications et redémarrez Hue comme indiqué dans l'exemple suivant :

```
sudo stop hue; sudo start hue
```

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.22.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.22.0 d'Amazon. Les modifications ont été apportées à la version 5.21.0.

Important

Depuis la EMR version 5.22.0 d'Amazon, Amazon EMR utilise AWS Signature version 4 exclusivement pour authentifier les demandes adressées à Amazon S3. EMR Les versions antérieures d'Amazon utilisent AWS la version 2 de Signature dans certains cas, sauf si les notes de publication indiquent que la version 4 de Signature est utilisée exclusivement. Pour plus d'informations, consultez les [sections Authentification des demandes \(AWS Signature version 4\)](#) et [Authentification des demandes \(AWS Signature version 2\)](#) dans le manuel Amazon Simple Storage Service Developer Guide.

Date de parution initiale : 20 mars 2019

Mises à niveau

- Flink 1.7.1
- HBase1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Connecteurs et pilotes :
 - Connecteur DynamoDB 4.8.0
 - Connecteur MariaDB 2.2.6
 - Pilote Amazon Redshift JDBC 1.2.20.1043

Nouvelles fonctionnalités

- Modification de la EBS configuration par défaut pour les types d'EC2instances avec stockage EBS uniquement. Lorsque vous créez un cluster à l'aide d'Amazon EMR version 5.22.0 ou ultérieure, la quantité de EBS stockage par défaut augmente en fonction de la taille de l'instance. En outre, nous avons réparti le stockage accru sur plusieurs volumes, ce qui a permis d'améliorer les performances. Si vous souhaitez utiliser une autre configuration de stockage d'EBSinstance, vous pouvez la spécifier lorsque vous créez un EMR cluster ou que vous ajoutez des nœuds à un cluster existant. Pour plus d'informations sur la quantité de stockage et le nombre de volumes alloués par défaut pour chaque type d'instance, consultez la section [EBSStockage par défaut pour les instances](#) dans le guide EMR de gestion Amazon.

Modifications, améliorations et problèmes résolus

- Spark
 - Ajout d'une nouvelle propriété de configuration pour Spark `onYARN,spark.yarn.executor.memoryOverheadFactor`. La valeur de cette propriété est un facteur d'échelle qui définit la valeur de la surcharge mémoire à un pourcentage de la mémoire de l'exécuteur, avec un minimum de 384 Mo. Si la surcharge mémoire est définie explicitement à l'aide `spark.yarn.executor.memoryOverhead`, cette propriété n'a aucun effet. La valeur par défaut est `0.1875`, ce qui représente 18,75 %. Cette valeur par défaut pour Amazon EMR laisse plus d'espace dans les YARN conteneurs pour la surcharge de mémoire de l'exécuteur que la valeur par défaut de 10 % définie en interne par Spark. La EMR valeur par

défaut d'Amazon de 18,75 % a montré empiriquement moins de défaillances liées à la mémoire dans les benchmarks -DS. TPC

- Rétroporté [SPARK-26316](#) pour améliorer les performances.
- Dans les EMR versions 5.19.0, 5.20.0 et 5.21.0 d'Amazon, les étiquettes des YARN nœuds sont stockées dans un répertoire. HDFS Dans certaines situations, cela entraîne des retards dans le démarrage des nœuds principaux, ce qui provoque le dépassement du délai du cluster et l'échec du lancement. À partir d'Amazon EMR 5.22.0, ce problème est résolu. YARN les étiquettes des nœuds sont stockées sur le disque local de chaque nœud du cluster, ce qui permet d'éviter toute dépendance vis-à-vis de HDFS.

Problèmes connus

- Hue (corrigé dans la EMR version 5.24.0 d'Amazon)
 - Hue exécuté sur Amazon EMR ne prend pas en charge Solr. À compter de la EMR version 5.20.0 d'Amazon, un problème de configuration incorrecte entraîne l'activation de Solr et un message d'erreur inoffensif similaire au suivant s'affiche :

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

Pour empêcher l'affichage du message d'erreur de Solr, procédez comme suit :

1. Connectez-vous à la ligne de commande du nœud principal à l'aide de SSH.
2. Utilisez un éditeur de texte pour ouvrir le fichier `hue.ini`. Par exemple :

```
sudo vim /etc/hue/conf/hue.ini
```

3. Recherchez le terme `appblacklist` et modifiez la ligne comme suit :

```
appblacklist = search
```

4. Enregistrez vos modifications et redémarrez Hue comme indiqué dans l'exemple suivant :

```
sudo stop hue; sudo start hue
```


- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.21.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.21.1 d'Amazon. Les modifications ont été apportées à la version 5.21.0.

Date de parution initiale : 18 juillet 2019

Modifications, améliorations et problèmes résolus

- Mise à jour de la version par défaut d'Amazon Linux AMI pour Amazon EMR afin d'inclure d'importantes mises à jour de sécurité du noyau Linux, notamment le problème de TCP SACK déni de service ([AWS-2019-005](#)).

Problèmes connus

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.21.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.21.0 d'Amazon. Les modifications ont été apportées à la version 5.20.0.

Date de parution initiale : 18 février 2019

Dernière mise à jour : 3 avril 2019

Mises à niveau

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1,11,479

Nouvelles fonctionnalités

- (3 avril 2019) Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier des classifications de configuration supplémentaires pour chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, utilisez la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).

Modifications, améliorations et problèmes résolus

- Zeppelin
 - Rétroporté [ZEPPELIN-3878](#).

Problèmes connus

- Hue (corrigé dans la EMR version 5.24.0 d'Amazon)
 - Hue exécuté sur Amazon EMR ne prend pas en charge Solr. À compter de la EMR version 5.20.0 d'Amazon, un problème de configuration incorrecte entraîne l'activation de Solr et un message d'erreur inoffensif similaire au suivant s'affiche :

```
Solr server could not be contacted properly:  
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
```

```
port=1978): Max retries exceeded with url: /solr/admin/info/system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

Pour empêcher l'affichage du message d'erreur de Solr, procédez comme suit :

1. Connectez-vous à la ligne de commande du nœud principal à l'aide de SSH.
2. Utilisez un éditeur de texte pour ouvrir le fichier `hue.ini`. Par exemple :

```
sudo vim /etc/hue/conf/hue.ini
```

3. Recherchez le terme `appblacklist` et modifiez la ligne comme suit :

```
appblacklist = search
```

4. Enregistrez vos modifications et redémarrez Hue comme indiqué dans l'exemple suivant :

```
sudo stop hue; sudo start hue
```

- Tez

- Ce problème a été résolu dans Amazon EMR 5.22.0.

Lorsque vous vous connectez à l'interface utilisateur de Tez sur `http://MasterDNS:8080/tez-ui` via une SSH connexion au nœud principal du cluster, l'erreur « L'opération de l'adaptateur a échoué - Le serveur Timeline (ATS) est hors de portée. Soit il est en panne, soit il CORS est « non activé » s'affiche, soit les tâches affichent « N/A » de manière inattendue.

Cela est dû au fait que l'interface utilisateur de Tez envoie des demandes au serveur YARN Timeline en utilisant `localhost` plutôt que le nom d'hôte du nœud principal. Pour contourner le problème, un script est disponible pour être exécuté en tant qu'action ou étape d'amorçage. Le script met à jour le nom d'hôte dans le fichier `Tez configs.env`. Pour plus d'informations et pour connaître l'emplacement du script, consultez les [Instructions sur l'amorçage](#).

- Dans les EMR versions 5.19.0, 5.20.0 et 5.21.0 d'Amazon, les étiquettes des YARN nœuds sont stockées dans un répertoire. HDFS Dans certaines situations, cela entraîne des retards dans le démarrage des nœuds principaux, ce qui provoque le dépassement du délai du cluster et l'échec du lancement. À partir d'Amazon EMR 5.22.0, ce problème est résolu. YARN Les étiquettes des nœuds sont stockées sur le disque local de chaque nœud du cluster, ce qui permet d'éviter toute dépendance vis-à-vis de HDFS.

- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSHen tant qu'hadooputilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```

Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.20.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.20.0 d'Amazon. Les modifications ont été apportées à la version 5.19.0.

Date de parution initiale : 18 décembre 2018

Dernière mise à jour : 22 janvier 2019

Mises à niveau

- Flink 1.6.2
- HBase1.4.8
- Hive 2.3.4
- Hue 4.3.0
- MXNet1.3.1
- Presto 0.214
- Spark 2.4.0
- TensorFlow 1,12,0
- Tez 0.9.1
- AWS SDK for Java 1,11,461

Nouvelles fonctionnalités

- (22 janvier 2019) Kerberos dans Amazon EMR a été amélioré pour prendre en charge l'authentification des principaux depuis un outil externe. KDC Cela permet de centraliser la gestion principale, car plusieurs clusters peuvent partager un seul cluster externeKDC. En outre, l'entité externe KDC peut bénéficier d'une confiance entre domaines avec un domaine Active Directory. Cela permet à tous les clusters d'authentifier les mandataires à partir d'Active Directory. Pour plus d'informations, consultez [Utiliser l'authentification Kerberos](#) dans le guide de EMRgestion Amazon.

Modifications, améliorations et problèmes résolus

- Amazon Linux par défaut AMI pour Amazon EMR
 - Le package Python3 a été mis à jour de python 3.4 à 3.6.
- Le EMRFS transmetteur optimisé pour S3
 - Le validateur EMRFS optimisé pour S3 est désormais activé par défaut, ce qui améliore les performances d'écriture. Pour de plus amples informations, veuillez consulter [Utilisez le transmetteur EMRFS optimisé pour S3](#).
- Hive
 - Rétroporté [HIVE-16686](#).
- Glue avec Spark et Hive

- Dans la version EMR 5.20.0 ou ultérieure, l'élagage parallèle des partitions est activé automatiquement pour Spark et Hive lorsque AWS Glue Data Catalog est utilisé comme métastore. Cette modification réduit considérablement le temps de planification des requêtes en exécutant plusieurs requêtes en parallèle pour récupérer des partitions. Le nombre total de segments pouvant être exécutés simultanément est compris entre 1 et 10. La valeur par défaut est 5, ce qui est recommandé. Vous pouvez le modifier en spécifiant la propriété `aws.glue.partition.num.segments` dans la classification de configuration `hive-site`. En cas de limitation, vous pouvez désactiver la fonctionnalité en remplaçant la valeur par 1. Pour en savoir plus, consultez [Structure d'un segment AWS Glue](#).

Problèmes connus

- Hue (corrigé dans la EMR version 5.24.0 d'Amazon)
- Hue exécuté sur Amazon EMR ne prend pas en charge Solr. À compter de la EMR version 5.20.0 d'Amazon, un problème de configuration incorrecte entraîne l'activation de Solr et un message d'erreur inoffensif similaire au suivant s'affiche :

```
Solr server could not be contacted properly:
HTTPConnectionPool('host=ip-xx-xx-xx-xx.ec2.internal',
port=1978): Max retries exceeded with url: /solr/admin/info/
system?user.name=hue&doAs=administrator&wt=json (Caused by
NewConnectionError(': Failed to establish a new connection: [Errno 111]
Connection refused',))
```

Pour empêcher l'affichage du message d'erreur de Solr, procédez comme suit :

1. Connectez-vous à la ligne de commande du nœud principal à l'aide de SSH.
2. Utilisez un éditeur de texte pour ouvrir le fichier `hue.ini`. Par exemple :

```
sudo vim /etc/hue/conf/hue.ini
```

3. Recherchez le terme `appblacklist` et modifiez la ligne comme suit :

```
appblacklist = search
```

4. Enregistrez vos modifications et redémarrez Hue comme indiqué dans l'exemple suivant :

```
sudo stop hue; sudo start hue
```

- Tez

- Ce problème a été résolu dans Amazon EMR 5.22.0.

Lorsque vous vous connectez à l'interface utilisateur de Tez sur `http://MasterDNS:8080/tez-ui` via une SSH connexion au nœud principal du cluster, l'erreur « L'opération de l'adaptateur a échoué - Le serveur Timeline (ATS) est hors de portée. Soit il est en panne, soit il CORS est « non activé » s'affiche, soit les tâches affichent « N/A » de manière inattendue.

Cela est dû au fait que l'interface utilisateur de Tez envoie des demandes au serveur YARN Timeline en utilisant `localhost` plutôt que le nom d'hôte du nœud principal. Pour contourner le problème, un script est disponible pour être exécuté en tant qu'action ou étape d'amorçage. Le script met à jour le nom d'hôte dans le fichier Tez `configs.env`. Pour plus d'informations et pour connaître l'emplacement du script, consultez les [Instructions sur l'amorçage](#).

- Dans les EMR versions 5.19.0, 5.20.0 et 5.21.0 d'Amazon, les étiquettes des YARN nœuds sont stockées dans un répertoire. HDFS Dans certaines situations, cela entraîne des retards dans le démarrage des nœuds principaux, ce qui provoque le dépassement du délai du cluster et l'échec du lancement. À partir d'Amazon EMR 5.22.0, ce problème est résolu. YARN Les étiquettes des nœuds sont stockées sur le disque local de chaque nœud du cluster, ce qui permet d'éviter toute dépendance vis-à-vis de HDFS.
- Problème connu dans les clusters dotés de plusieurs nœuds primaires et d'une authentification Kerberos

Si vous exécutez des clusters avec plusieurs nœuds principaux et une authentification Kerberos dans les EMR versions 5.20.0 et ultérieures d'Amazon, vous risquez de rencontrer des problèmes avec les opérations de cluster, telles que la réduction ou la soumission par étapes, une fois le cluster en cours d'exécution pendant un certain temps. La durée dépend de la période de validité du ticket Kerberos que vous avez définie. Le problème de réduction d'échelle a un impact à la fois sur la réduction d'échelle automatique et sur les demandes de réduction d'échelle explicites que vous avez soumises. D'autres opérations de cluster peuvent également être affectées.

Solution :

- SSH en tant qu'hadoop utilisateur du nœud principal du EMR cluster comportant plusieurs nœuds principaux.
- Exécutez la commande suivante pour renouveler le ticket Kerberos pour l'utilisateur hadoop.

```
kinit -kt <keytab_file> <principal>
```


Généralement, le fichier keytab se trouve dans `/etc/hadoop.keytab` et le principal se présente sous la forme de `hadoop/<hostname>@<REALM>`.

Note

Cette solution de contournement sera effective pendant toute la durée de validité du ticket Kerberos. Cette durée est de 10 heures par défaut, mais peut être configurée par vos paramètres Kerberos. Vous devez exécuter à nouveau la commande ci-dessus une fois le ticket Kerberos expiré.

Version 5.19.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.19.0 d'Amazon. Les modifications ont été apportées à la version 5.18.0.

Date de parution initiale : 7 novembre 2018

Date de la dernière mise à jour : 19 novembre 2018

Mises à niveau

- Hadoop 2.8.5
- Flink 1.6.1
- JupyterHub 0,9.4
- MXNet1.3.0
- Presto 0.212
- TensorFlow 1,11,0
- Zookeeper 3.4.13
- AWS SDK for Java 1,1,433

Nouvelles fonctionnalités

- (19 novembre 2018) EMR Notebooks est un environnement géré basé sur Jupyter Notebook. Il prend en charge les noyaux magiques Spark pour PySpark, SparkSQL, Spark R et Scala. EMR Les blocs-notes peuvent être utilisés avec des clusters créés à l'aide des EMR versions 5.18.0 et

ultérieures d'Amazon. Pour plus d'informations, consultez la section [Utilisation EMR des blocs-notes](#) dans le guide EMR de gestion Amazon.

- Le valideur EMRFS optimisé pour S3 est disponible lors de l'écriture de fichiers Parquet à l'aide de Spark et. EMRFS Ce validateur améliore les performances d'écriture. Pour de plus amples informations, veuillez consulter [Utilisez le transmetteur EMRFS optimisé pour S3](#).

Modifications, améliorations et problèmes résolus

- YARN
 - Modification de la logique qui limite le processus principal de l'application à l'exécution sur les nœuds principaux. Cette fonctionnalité utilise désormais la fonctionnalité et les propriétés des étiquettes de YARN nœuds dans les classifications `capacity-scheduler` de configuration `yarn-site` et. Pour plus d'informations, veuillez consulter <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.
- Amazon Linux par défaut AMI pour Amazon EMR
 - `ruby18`, `php56` et `gcc48` ne sont plus installés par défaut. Ils peuvent être installés si vous le souhaitez à l'aide de `yum`.
 - Le gem `ruby aws-sdk` n'est plus installé par défaut. Il peut être installé en utilisant `gem install aws-sdk`, si vous le souhaitez. Des composants spécifiques peuvent également être installés. Par exemple, `gem install aws-sdk-s3`.

Problèmes connus

- EMRBlocs-notes : dans certains cas, lorsque plusieurs éditeurs de blocs-notes sont ouverts, l'éditeur de bloc-notes peut sembler incapable de se connecter au cluster. Dans ce cas, effacez les cookies du navigateur, puis rouvrez les éditeurs de bloc-notes.
- CloudWatch ContainerPending Mise à l'échelle métrique et automatique — (Corrigé dans la version 5.20.0) Amazon EMR peut émettre une valeur négative pour `ContainerPending`. Si `ContainerPending` est utilisée dans une règle de mise à l'échelle automatique, la mise à l'échelle automatique ne se comporte pas comme prévu. Évitez d'utiliser `ContainerPending` avec la mise à l'échelle automatique.
- Dans les EMR versions 5.19.0, 5.20.0 et 5.21.0 d'Amazon, les étiquettes des YARN nœuds sont stockées dans un répertoire. HDFS Dans certaines situations, cela entraîne des retards dans le démarrage des nœuds principaux, ce qui provoque le dépassement du délai du cluster et l'échec du lancement. À partir d'Amazon EMR 5.22.0, ce problème est résolu. YARN Les étiquettes des

nœuds sont stockées sur le disque local de chaque nœud du cluster, ce qui permet d'éviter toute dépendance vis-à-vis deHDFS.

Version 5.18.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.18.0 d'Amazon. Les modifications ont été apportées à la version 5.17.0.

Date de parution initiale : 24 octobre 2018

Mises à niveau

- Flink 1.6.0
- HBase1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

Nouvelles fonctionnalités

- À partir d'Amazon EMR 5.18.0, vous pouvez utiliser le référentiel d'EMRartefacts Amazon pour créer votre code de travail en fonction des versions exactes des bibliothèques et des dépendances disponibles avec des versions spécifiques d'Amazon. EMR Pour de plus amples informations, veuillez consulter [Vérification des dépendances à l'aide du référentiel d'artefacts d'Amazon EMR](#).

Modifications, améliorations et problèmes résolus

- Hive
 - Ajout de la prise en charge de S3 Select. Pour de plus amples informations, veuillez consulter [Utilisation de S3 Select avec Hive pour améliorer les performances](#).
- Presto
 - Ajout de la prise en charge de [S3 Select](#) Pushdown. Pour de plus amples informations, veuillez consulter [Utilisation de S3 Select Pushdown avec Presto pour améliorer les performances](#).
- Spark

- La configuration par défaut de log4j pour Spark a été modifiée pour enregistrer les journaux des conteneurs toutes les heures pour les tâches de streaming Spark. Cela permet d'éviter la suppression des journaux pour les tâches de streaming Spark de longue durée.

Version 5.17.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.17.1 d'Amazon. Les modifications ont été apportées à la version 5.17.0.

Date de parution initiale : 18 juillet 2019

Modifications, améliorations et problèmes résolus

- Mise à jour de la version par défaut d'Amazon Linux AMI pour Amazon EMR afin d'inclure d'importantes mises à jour de sécurité du noyau Linux, notamment le problème de TCP SACK déni de service ([AWS-2019-005](#)).

Version 5.17.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.17.0 d'Amazon. Les modifications ont été apportées à la version 5.16.0.

Date de parution initiale : 30 août 2018

Mises à niveau

- Flink 1.5.2
- HBase1.4.6
- Presto 0.206

Nouvelles fonctionnalités

- Ajout de la prise en charge de Tensorflow. Pour de plus amples informations, veuillez consulter [TensorFlow](#).

Modifications, améliorations et problèmes résolus

- JupyterHub

- Ajout de la prise en charge de la persistance des blocs-notes dans Amazon S3. Pour de plus amples informations, veuillez consulter [Configuration de la persistance pour les blocs-notes dans Amazon S3](#).
- Spark
 - Ajout de la prise en charge de [S3 Select](#). Pour de plus amples informations, veuillez consulter [Utilisation de S3 Select avec Spark pour améliorer les performances des requêtes](#).
- Résolution des problèmes liés aux métriques Cloudwatch et à la fonctionnalité de dimensionnement automatique dans les EMR versions 5.14.0, 5.15.0 ou 5.16.0 d'Amazon.

Problèmes connus

- Lorsque vous créez un cluster activé pour Kerberos avec Livy installé, Livy échoue avec un message d'erreur indiquant que l'authentification simple n'est pas activée. Le redémarrage du serveur Livy résout le problème. Pour contourner le problème, ajoutez une étape lors de la création du cluster qui exécute `sudo restart livy-server` sur le nœud primaire.
- Si vous utilisez un Amazon Linux personnalisé AMI basé sur un Amazon Linux AMI dont la date de création est le 11/08/2018, le serveur Oozie ne démarre pas. Si vous utilisez Oozie, créez une personnalisation AMI basée sur un AMI identifiant Amazon Linux avec une date de création différente. Vous pouvez utiliser la AWS CLI commande suivante pour renvoyer une liste d'images IDs pour tous les HVM Amazon Linux AMIs avec une version 2018.03, ainsi que la date de sortie, afin de pouvoir choisir un Amazon Linux approprié AMI comme base. MyRegion Remplacez-le par l'identifiant de votre région, tel que us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`][[?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

Version 5.16.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.16.0 d'Amazon. Les modifications ont été apportées à la version 5.15.0.

Date de parution initiale : 19 juillet 2018

Mises à niveau

- Hadoop 2.8.4

- Flink 1.5.0
- Livy 0.5.0
- MXNet1.2.0
- Phoenix 4.14.0
- Presto 0.203
- Spark 2.3.1
- AWS SDK for Java 1,1336
- CUDA9,2
- Pilote Redshift 1.2.15.1025 JDBC

Modifications, améliorations et problèmes résolus

- HBase
 - [Rétroporté -20723 HBASE](#)
- Presto
 - Modifications de configuration pour prendre en charge LDAP l'authentification. Pour de plus amples informations, veuillez consulter [Utilisation de LDAP l'authentification pour Presto sur Amazon EMR](#).
- Spark
 - [Apache Spark version 2.3.1, disponible à partir de la EMR version 5.16.0 d'Amazon, adresses CVE-2018-8024 et -2018-1334. CVE](#) Nous vous recommandons de migrer les versions antérieures de Spark vers la version 2.3.1 ou ultérieure.

Problèmes connus

- Cette version ne prend pas en charge les types d'instance c1.medium ou m1.small. Les clusters utilisant l'un ou l'autre de ces types d'instances ne démarrent pas. Pour contourner le problème, spécifiez un type d'instance différent ou utilisez une version différente.
- Lorsque vous créez un cluster activé pour Kerberos avec Livy installé, Livy échoue avec un message d'erreur indiquant que l'authentification simple n'est pas activée. Le redémarrage du serveur Livy résout le problème. Pour contourner le problème, ajoutez une étape lors de la création du cluster qui exécute `sudo restart livy-server` sur le nœud primaire.

- Après le redémarrage du nœud principal ou le redémarrage du contrôleur d'instance, les CloudWatch métriques ne seront pas collectées et la fonctionnalité de dimensionnement automatique ne sera pas disponible dans les EMR versions 5.14.0, 5.15.0 ou 5.16.0 d'Amazon. Ce problème est résolu dans Amazon EMR 5.17.0.

Version 5.15.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.15.0 d'Amazon. Les modifications ont été apportées à la version 5.14.0.

Date de parution initiale : 21 juin 2018

Mises à niveau

- Mise à niveau HBase vers la version 1.4.4
- Mise à niveau de Hive vers la version 2.3.3
- Mise à niveau de Hue vers la version 4.2.0
- Mise à niveau d'Oozie vers la version 5.0.0
- Mise à niveau de Zookeeper vers la version 3.4.12
- Mise à niveau AWS SDK vers la version 1.11.333

Modifications, améliorations et problèmes résolus

- Hive
 - [Rétroporté -18069 HIVE](#)
- Hue
 - Hue a été mis à jour pour s'authentifier correctement auprès de Livy lorsque Kerberos est activé. Livy est désormais pris en charge lors de l'utilisation de Kerberos avec Amazon. EMR
- JupyterHub
 - Mis à jour JupyterHub afin qu'Amazon EMR installe les bibliothèques LDAP clientes par défaut.
 - Correction d'une erreur dans le script qui génère des certificats auto-signés.

Problèmes connus

- Cette version ne prend pas en charge les types d'instance c1.medium ou m1.small. Les clusters utilisant l'un ou l'autre de ces types d'instances ne démarrent pas. Pour contourner le problème, spécifiez un type d'instance différent ou utilisez une version différente.
- Après le redémarrage du nœud principal ou le redémarrage du contrôleur d'instance, les CloudWatch métriques ne seront pas collectées et la fonctionnalité de dimensionnement automatique ne sera pas disponible dans les EMR versions 5.14.0, 5.15.0 ou 5.16.0 d'Amazon. Ce problème est résolu dans Amazon EMR 5.17.0.

Version 5.14.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.14.1 d'Amazon. Les modifications ont été apportées à la version 5.14.0.

Date de parution initiale : 17 octobre 2018

Mise à jour de la valeur par défaut AMI pour Amazon EMR afin de corriger les failles de sécurité potentielles.

Version 5.14.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.14.0 d'Amazon. Les modifications ont été apportées à la version 5.13.0.

Date de parution initiale : 4 juin 2018

Mises à niveau

- Mise à niveau d'Apache Flink vers la version 1.4.2
- Mise à niveau d'Apache MXnet vers la version 1.1.0
- Mise à niveau d'Apache Sqoop vers la version 1.4.7

Nouvelles fonctionnalités

- JupyterHub Support ajouté. Pour de plus amples informations, veuillez consulter [JupyterHub](#).

Modifications, améliorations et problèmes résolus

- EMRFS
 - La userAgent chaîne des demandes adressées à Amazon S3 a été mise à jour pour contenir les informations d'utilisateur et de groupe du principal invocateur. Cela peut être utilisé avec AWS CloudTrail les journaux pour un suivi plus complet des demandes.
- HBase
 - Inclus [HBASE-20447](#), qui résout un problème susceptible de provoquer des problèmes de cache, en particulier dans le cas de régions divisées.
- MXnet
 - Ajout des bibliothèques OpenCV.
- Spark
 - Lorsque Spark écrit des fichiers Parquet sur un emplacement Amazon S3 en utilisant EMRFS, l'FileOutputCommitter algorithme a été mis à jour pour utiliser la version 2 au lieu de la version 1. Cela réduit le nombre de renommages, ce qui améliore les performances de l'application. Cette modification n'affecte pas :
 - Les applications autres que Spark.
 - Applications qui écrivent sur d'autres systèmes de fichiers, tels que HDFS (qui utilisent toujours la version 1 de FileOutputCommitter).
 - Applications utilisant d'autres formats de sortie, tels que le texte ou le csv, qui utilisent déjà l'écriture EMRFS directe.

Problèmes connus

- JupyterHub
 - L'utilisation de classifications de configuration pour configurer JupyterHub des blocs-notes Jupyter individuels lors de la création d'un cluster n'est pas prise en charge. Modifiez manuellement le fichier `jupyterhub_config.py` et les fichiers `jupyter_notebook_config.py` pour chaque utilisateur. Pour de plus amples informations, veuillez consulter [Configuration JupyterHub](#).
 - JupyterHub ne démarre pas sur des clusters au sein d'un sous-réseau privé, ce qui entraîne un échec du message `error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'`. Ce problème est dû à une erreur dans le script qui génère

des certificats auto-signés. Utilisez la solution suivante pour générer des certificats auto-signés. Toutes les commandes sont exécutées lorsque vous êtes connecté au nœud primaire.

1. Copiez le script de génération de certificat du conteneur vers le nœud primaire :

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Utilisez un éditeur de texte pour modifier la ligne 23 afin de changer le nom d'hôte public en nom d'hôte local, comme indiqué ci-dessous :

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Exécutez le script pour générer des certificats auto-signés :

```
sudo bash ./gen_self_signed_cert.sh
```

4. Déplacez les fichiers de certificat générés par le script vers le répertoire `/etc/jupyter/conf/` :

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

Vous pouvez vérifier que `tail` le `jupyter.log` fichier JupyterHub a redémarré et qu'il renvoie un code de réponse 200. Par exemple :

```
tail -f /var/log/jupyter/jupyter.log
```

Vous devriez obtenir une réponse similaire à la suivante :

```
# [I 2018-06-14 18:56:51.356 JupyterHub app:1581] JupyterHub is now running at
https://:9443/
# 19:01:51.359 - info: [ConfigProxy] 200 GET /api/routes
```

- Après le redémarrage du nœud principal ou le redémarrage du contrôleur d'instance, les CloudWatch métriques ne seront pas collectées et la fonctionnalité de dimensionnement automatique ne sera pas disponible dans les EMR versions 5.14.0, 5.15.0 ou 5.16.0 d'Amazon. Ce problème est résolu dans Amazon EMR 5.17.0.

Version 5.13.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.13.0 d'Amazon. Les modifications ont été apportées à la version 5.12.0.

Mises à niveau

- Mise à niveau de Spark vers la version 2.3.0
- Mise à niveau HBase vers la version 1.4.2
- Mise à niveau vers de Presto vers la version 0.194
- Mise à niveau AWS SDK for Java vers la version 1.11.297

Modifications, améliorations et problèmes résolus

- Hive
 - Rétroporté [HIVE-15436](#). Hive améliorée pour ne APIs renvoyer que les vues.

Problèmes connus

- MXNet ne possède actuellement pas de bibliothèques OpenCV.

Version 5.12.2

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.12.2 d'Amazon. Les modifications ont été apportées à la version 5.12.1.

Date de parution initiale : 29 août 2018

Modifications, améliorations et problèmes résolus

- Cette version corrige une vulnérabilité de sécurité potentielle.

Version 5.12.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.12.1 d'Amazon. Les modifications ont été apportées à la version 5.12.0.

Date de parution initiale : 29 mars 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités potentielles.

Version 5.12.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.12.0 d'Amazon. Les modifications ont été apportées à la version 5.11.1.


Mises à niveau

- AWS SDK pour Java 1.11.238 ⇒ 1.11.267. Pour plus d'informations, consultez la section [AWS SDK for Java Change Log](#) on GitHub.
- Hadoop 2.7.3 ⇒ 2.8.3. Pour plus d'informations, consultez [Versions d'Apache Hadoop](#).
- Flink 1.3.2 ⇒ 1.4.0. Pour plus d'informations, consultez [Annonce de publication d'Apache Flink 1.4.0](#).
- HBase 1.3.1 ⇒ 1.4.0. Pour plus d'informations, consultez l'[annonce HBase de sortie](#).
- Hue 4.0.1 ⇒ 4.1.0. Pour plus d'informations, veuillez consulter les [Notes de mise à jour](#).
- MxNet 0,12,0 ⇒ 10,0. Pour plus d'informations, consultez la section [MXNet Change Log](#) on GitHub.
- Presto 0.187 ⇒ 0.188. Pour plus d'informations, veuillez consulter les [Notes de mise à jour](#).

Modifications, améliorations et problèmes résolus

- Hadoop
 - La propriété `yarn.resourcemanager.decommissioning.timeout` a été remplacée par `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`. Vous pouvez utiliser cette propriété pour personnaliser la réduction de la taille du cluster. Pour plus d'informations, consultez [Cluster Scale-Down dans le](#) guide de EMR gestion Amazon.
 - Hadoop CLI a ajouté l'option à la commande `cp` (copy), qui spécifie la copie directe. Vous pouvez l'utiliser pour éviter de créer un fichier `.COPYING` intermédiaire, ce qui accélère la copie de données entre Amazon S3. Pour plus d'informations, consultez [HADOOP-12384](#).
- Pig

- Ajout de la classification de configuration `pig-env`, qui simplifie la configuration des propriétés de l'environnement Pig. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).
- Presto
 - Ajout de la classification de configuration, `presto-connector-redshift` que vous pouvez utiliser pour configurer des valeurs dans le fichier de configuration `redshift.properties` de Presto. Pour plus d'informations, consultez [Connecteur Redshift](#) dans la documentation Presto, et [Configuration des applications](#).
 - La prise en charge de Presto pour EMRFS a été ajoutée et constitue la configuration par défaut. Les EMR versions précédentes d'Amazon utilisaient `PrestOS3FileSystem`, qui était la seule option. Pour de plus amples informations, veuillez consulter [EMRFS configuration de PrestOS3 FileSystem](#).

 Note

Si vous interrogez des données sous-jacentes dans Amazon S3 avec Amazon EMR version 5.12.0, des erreurs Presto peuvent se produire. Cela est dû au fait que Presto ne parvient pas à récupérer les valeurs de classification de configuration depuis `emrfs-site.xml`. Pour contourner le problème, créez un sous-répertoire `emrfs` sous `usr/lib/presto/plugin/hive-hadoop2/` et un lien symbolique dans `usr/lib/presto/plugin/hive-hadoop2/emrfs` vers le fichier `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` existant. Redémarrez ensuite le processus `presto-server` (`sudo presto-server stop` suivi de `sudo presto-server start`).

- Spark
 - Rétroporté [SPARK-22036 : la BigDecimal multiplication renvoie parfois la valeur nulle](#).

Problèmes connus

- MXNet n'inclut pas les bibliothèques OpenCV.
- SparkR n'est pas disponible pour les clusters créés à l'aide d'un custom car AMI R n'est pas installé par défaut sur les nœuds du cluster.

Version 5.11.3

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.11.3 d'Amazon. Les modifications ont été apportées à la version 5.11.2.

Date de parution initiale : 18 juillet 2019

Modifications, améliorations et problèmes résolus

- Mise à jour de la version par défaut d'Amazon Linux AMI pour Amazon EMR afin d'inclure d'importantes mises à jour de sécurité du noyau Linux, notamment le problème de TCP SACK déni de service ([AWS-2019-005](#)).

Version 5.11.2

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.11.2 d'Amazon. Les modifications ont été apportées à la version 5.11.1.

Date de parution initiale : 29 août 2018

Modifications, améliorations et problèmes résolus

- Cette version corrige une vulnérabilité de sécurité potentielle.

Version 5.11.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.11.1 d'Amazon. Les modifications concernent la version EMR 5.11.0 d'Amazon.

Date de parution initiale : 22 janvier 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités associées à l'exécution spéculative (CVE-2017-5715, -2017-5753 et CVE -2017-5754). CVE Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Problèmes connus

- MXNet n'inclut pas les bibliothèques OpenCV.
- Hive 2.3.2 définit `hive.compute.query.using.stats=true` par défaut. Cela entraîne des requêtes pour obtenir des données à partir de statistiques existantes plutôt que directement à partir des données, ce qui peut être déroutant. Par exemple, si vous avez une table avec `hive.compute.query.using.stats=true` et que vous téléchargez de nouveaux fichiers vers la table LOCATION, l'exécution d'une demande `SELECT COUNT(*)` sur la table renvoie le nombre des statistiques, plutôt que de récupérer les lignes ajoutées.

Pour contourner ce problème, utilisez la commande `ANALYZE TABLE` pour collecter de nouvelles statistiques, ou définissez `hive.compute.query.using.stats=false`. Pour en savoir plus, consultez [Statistiques dans Hive](#) dans la documentation Apache Hive.

Version 5.11.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.11.0 d'Amazon. Les modifications concernent la EMR version 5.10.0 d'Amazon.

Mises à niveau

Les applications et les composants suivants ont été mis à niveau dans cette édition pour inclure les versions suivantes :

- Hive 2.3.2
- Spark 2.2.1
- SDK pour Java 1.11.238

Nouvelles fonctionnalités

- Spark
 - Ajout du paramètre `spark.decommissioning.timeout.threshold`, qui améliore le comportement de mise hors service de Spark lors de l'utilisation d'instances Spot. Pour de plus amples informations, veuillez consulter [Configuration du comportement de mise hors service du nœud](#).
 - Ajout du `aws-sagemaker-spark-sdk` composant à Spark, qui installe Amazon SageMaker Spark et les dépendances associées pour l'intégration de Spark à [Amazon SageMaker](#). Vous

pouvez utiliser Amazon SageMaker Spark pour créer des pipelines d'apprentissage automatique (ML) Spark à l'aide d'Amazon SageMaker Stages. Pour plus d'informations, consultez le [fichier readme de SageMaker Spark](#) sur GitHub et [l'utilisation d'Apache Spark avec Amazon SageMaker](#) dans le manuel Amazon SageMaker Developer Guide.

Problèmes connus

- MXNet n'inclut pas les bibliothèques OpenCV.
- Hive 2.3.2 définit `hive.compute.query.using.stats=true` par défaut. Cela entraîne des requêtes pour obtenir des données à partir de statistiques existantes plutôt que directement à partir des données, ce qui peut être déroutant. Par exemple, si vous avez une table avec `hive.compute.query.using.stats=true` et que vous téléchargez de nouveaux fichiers vers la table LOCATION, l'exécution d'une demande `SELECT COUNT(*)` sur la table renvoie le nombre des statistiques, plutôt que de récupérer les lignes ajoutées.

Pour contourner ce problème, utilisez la commande `ANALYZE TABLE` pour collecter de nouvelles statistiques, ou définissez `hive.compute.query.using.stats=false`. Pour en savoir plus, consultez [Statistiques dans Hive](#) dans la documentation Apache Hive.

Version 5.10.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.10.0 d'Amazon. Les modifications concernent la EMR version 5.9.0 d'Amazon.

Mises à niveau

Les applications et les composants suivants ont été mis à niveau dans cette édition pour inclure les versions suivantes :

- AWS SDK for Java 1,11,221
- Hive 2.3.1
- Presto 0.187

Nouvelles fonctionnalités

- Ajout de la prise en charge de l'authentification Kerberos. Pour plus d'informations, consultez [Utiliser l'authentification Kerberos](#) dans le guide de gestion Amazon EMR

- Ajout de la prise en charge IAM des rôles pour les EMRFS demandes adressées à Amazon S3. Pour plus d'informations, consultez [Configurer IAM les rôles pour les EMRFS demandes adressées à Amazon S3](#) dans le guide EMR de gestion Amazon.
- Ajout de la prise en charge GPU des types d'instances basés sur P2 et P3. Pour plus d'informations, consultez les sections Instances [Amazon EC2 P2 et instances Amazon EC2 P3](#). NVIDIAle pilote 384.81 et CUDA le pilote 9.0.176 sont installés par défaut sur ces types d'instances.
- Ajout de la prise en charge de [Apache MXNet](#).

Modifications, améliorations et problèmes résolus

- Presto
 - Ajout de la prise en charge de l'utilisation du catalogue de données AWS Glue comme métastore Hive par défaut. Pour plus d'informations, consultez la section [Utilisation de Presto avec le catalogue de données AWS Glue](#).
 - Ajout de la prise en charge des [fonctions géospatiales](#).
 - Ajout de la prise en charge du [déversement sur le disque](#) pour les jointures.
 - Ajout de la prise en charge du [connecteur Redshift](#).
- Spark
 - Rétroporté [SPARK-20640](#), ce qui rend le délai d'expiration du RPC et les nouvelles tentatives d'enregistrement aléatoire configurables à l'aide des propriétés et des propriétés. `spark.shuffle.registration.timeout`
`spark.shuffle.registration.maxAttempts`
 - Rétroporté [SPARK-21549](#), qui corrige une erreur survenue lors de l'écriture personnalisée dans des emplacements non OutputFormat localisés. HDFS
- Rétroportage de [Hadoop-13270](#)
- Les bibliothèques Numpy, Scipy et Matplotlib ont été supprimées de la base Amazon. EMR AMI Si vous avez besoin de ces bibliothèques pour votre application, elles sont disponibles dans le référentiel d'applications. Vous pouvez utiliser une action d'amorçage pour les installer sur tous les nœuds à l'aide de `yum install`.
- La EMR base Amazon AMI ne contient plus de RPM packages d'applications, de sorte que les RPM packages ne sont plus présents sur les nœuds du cluster. Custom AMIs et la EMR base Amazon AMI font désormais référence au référentiel de RPM packages dans Amazon S3.

- En raison de l'introduction de la facturation à la seconde sur AmazonEC2, le comportement de réduction par défaut est désormais de terminer la tâche à la fin de la tâche plutôt que de terminer à l'heure de l'instance. Pour plus d'informations, consultez [Configuration de la réduction de la capacité des clusters](#).

Problèmes connus

- MXNet n'inclut pas les bibliothèques OpenCV.
- Hive 2.3.1 définit `hive.compute.query.using.stats=true` par défaut. Cela entraîne des requêtes pour obtenir des données à partir de statistiques existantes plutôt que directement à partir des données, ce qui peut être déroutant. Par exemple, si vous avez une table avec `hive.compute.query.using.stats=true` et que vous téléchargez de nouveaux fichiers vers la table LOCATION, l'exécution d'une demande `SELECT COUNT(*)` sur la table renvoie le nombre des statistiques, plutôt que de récupérer les lignes ajoutées.

Pour contourner ce problème, utilisez la commande `ANALYZE TABLE` pour collecter de nouvelles statistiques, ou définissez `hive.compute.query.using.stats=false`. Pour en savoir plus, consultez [Statistiques dans Hive](#) dans la documentation Apache Hive.

Version 5.9.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.9.0 d'Amazon. Les modifications concernent la version EMR 5.8.0 d'Amazon.

Date de parution : 5 octobre 2017

Dernière date de mise à jour de la fonctionnalité : 12 octobre 2017

Mises à niveau

Les applications et les composants suivants ont été mis à niveau dans cette édition pour inclure les versions suivantes :

- AWS SDK for Java version 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0

- Presto 0.184

Nouvelles fonctionnalités

- Ajout de la prise en charge de Livy (version 0.4.0-incubating). Pour de plus amples informations, veuillez consulter [Apache Livy](#).
- Ajout de la prise en charge de Hue Notebook pour Spark.
- Ajout de la prise en charge des EC2 instances Amazon de la série i3 (12 octobre 2017).

Modifications, améliorations et problèmes résolus

- Spark
 - Ajout d'un nouvel ensemble de fonctionnalités qui permet de s'assurer que Spark gère plus élégamment la terminaison des nœuds suite à un redimensionnement manuel ou à une demande de stratégie de dimensionnement automatique. Pour de plus amples informations, veuillez consulter [Configuration du comportement de mise hors service du nœud](#).
 - SSL est utilisé au lieu de 3 DES pour le chiffrement en transit du service de transfert par blocs, ce qui améliore les performances lors de l'utilisation des types d'EC2 instances Amazon avec AES - NI.
 - Rétroporté [SPARK-21494](#).
- Zeppelin
 - Rétroporté [ZEPPELIN-2377](#).
- HBase
 - Ajout du correctif [HBASE-18533](#), qui autorise des valeurs supplémentaires pour la configuration à l'aide HBase BucketCache de la classification de hbase-site configuration.
- Hue
 - Ajout de la prise en charge du catalogue de données AWS Glue pour l'éditeur de requêtes Hive dans Hue.
 - Par défaut, les superutilisateurs de Hue peuvent accéder à tous les fichiers auxquels les EMR IAM rôles Amazon sont autorisés à accéder. Les utilisateurs nouvellement créés n'ont pas automatiquement les autorisations d'accéder à l'explorateur de fichiers Amazon S3 et doivent avoir les autorisations `filebrowser.s3_access` activées pour leur groupe.
 - Résolution d'un problème en raison duquel les JSON données sous-jacentes créées à l'aide de AWS Glue Data Catalog étaient inaccessibles.

Problèmes connus

- Le lancement du cluster échoue lorsque toutes les applications sont installées et que la taille du volume EBS racine Amazon par défaut n'est pas modifiée. Pour contourner le problème, utilisez la `aws emr create-cluster` commande du AWS CLI et spécifiez un `--ebs-root-volume-size` paramètre plus grand.
- Hive 2.3.0 définit `hive.compute.query.using.stats=true` par défaut. Cela entraîne des requêtes pour obtenir des données à partir de statistiques existantes plutôt que directement à partir des données, ce qui peut être déroutant. Par exemple, si vous avez une table avec `hive.compute.query.using.stats=true` et que vous téléchargez de nouveaux fichiers vers la table LOCATION, l'exécution d'une demande `SELECT COUNT(*)` sur la table renvoie le nombre des statistiques, plutôt que de récupérer les lignes ajoutées.

Pour contourner ce problème, utilisez la commande `ANALYZE TABLE` pour collecter de nouvelles statistiques, ou définissez `hive.compute.query.using.stats=false`. Pour en savoir plus, consultez [Statistiques dans Hive](#) dans la documentation Apache Hive.

Version 5.8.2

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.8.2 d'Amazon. Les modifications ont été apportées à la version 5.8.1.

Date de parution initiale : 29 mars 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités potentielles.

Version 5.8.1

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.8.1 d'Amazon. Les modifications concernent la version EMR 5.8.0 d'Amazon.

Date de parution initiale : 22 janvier 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de default Amazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités associées à l'exécution spéculative (CVE-2017-5715, -2017-5753 et CVE-2017-5754). Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Version 5.8.0

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.8.0 d'Amazon. Les modifications concernent la EMR version 5.7.0 d'Amazon.

Date de parution initiale : 10 août 2017

Dernière date de mise à jour de la fonctionnalité : 25 septembre 2017

Mises à niveau

Les applications et les composants suivants ont été mis à jour dans cette version pour inclure les versions suivantes :

- AWS SDK 1,1,1160
- Flink 1.3.1
- Hive 2.3.0. Pour plus d'informations, consultez [Notes de mise à jour](#) sur le site Apache Hive.
- Spark 2.2.0. Pour plus d'informations, consultez [Notes de mise à jour](#) sur le site Apache Spark.

Nouvelles fonctionnalités

- Ajout de la prise en charge de l'affichage de l'historique de l'application (25 septembre 2017). Pour plus d'informations, consultez la section [Affichage de l'historique des applications](#) dans le guide EMR de gestion Amazon.

Modifications, améliorations et problèmes résolus

- Intégration avec AWS Glue Data Catalog
 - Hive et Spark ont désormais la possibilité SQL d'utiliser AWS Glue Data Catalog comme magasin de métadonnées Hive. Pour plus d'informations, consultez [Utiliser le catalogue de](#)

[données AWS Glue comme métastore pour Hive](#) et [Utiliser le catalogue de données AWS Glue comme métastore pour Spark SQL](#).

- Ajout de l'historique des applications aux détails du cluster, ce qui vous permet de consulter les données historiques des YARN applications et des détails supplémentaires pour les applications Spark. Pour plus d'informations, consultez la section [Afficher l'historique des applications](#) dans le guide EMR de gestion Amazon.
- Oozie
 - Rétroporté [OOZIE-2748](#).
- Hue
 - [Rétroporté -5859 HUE](#)
- HBase
 - Ajout d'un correctif pour exposer l'heure de démarrage du serveur HBase principal via Java Management Extensions (JMX) à l'aide de `getMasterInitializedTime`.
 - Ajout d'un correctif qui améliore la date de début du cluster.

Problèmes connus

- Le lancement du cluster échoue lorsque toutes les applications sont installées et que la taille du volume EBS racine Amazon par défaut n'est pas modifiée. Pour contourner le problème, utilisez la `aws emr create-cluster` commande du AWS CLI et spécifiez un `--ebs-root-volume-size` paramètre plus grand.
- Hive 2.3.0 définit `hive.compute.query.using.stats=true` par défaut. Cela entraîne des requêtes pour obtenir des données à partir de statistiques existantes plutôt que directement à partir des données, ce qui peut être déroutant. Par exemple, si vous avez une table avec `hive.compute.query.using.stats=true` et que vous téléchargez de nouveaux fichiers vers la table `LOCATION`, l'exécution d'une demande `SELECT COUNT(*)` sur la table renvoie le nombre des statistiques, plutôt que de récupérer les lignes ajoutées.

Pour contourner ce problème, utilisez la commande `ANALYZE TABLE` pour collecter de nouvelles statistiques, ou définissez `hive.compute.query.using.stats=false`. Pour en savoir plus, consultez [Statistiques dans Hive](#) dans la documentation Apache Hive.

- Spark – Il existe un problème de fuite du gestionnaire de fichiers avec le démon `apppusher` lors de l'utilisation de Spark, qui peut apparaître après plusieurs heures ou jours pour une tâche Spark de longue durée. Pour corriger le problème, connectez-vous au nœud maître et tapez `sudo /etc/`

`init.d/appusher stop`. Cela arrête le daemon `appusher`, qu'Amazon EMR redémarrera automatiquement.

- Historique de l'application
 - Les données d'historique pour les exécuteurs Spark inactifs ne sont pas disponibles.
 - L'historique de l'application n'est pas disponible pour les clusters qui utilisent une configuration de sécurité pour permettre le chiffrement à la volée.

Version 5.7.0

Les notes de mise à jour suivantes incluent des informations relatives à la version EMR 5.7.0 d'Amazon. Les modifications concernent la EMR version 5.6.0 d'Amazon.

Date de parution : 13 juillet 2017

Mises à niveau

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

Nouvelles fonctionnalités

- Ajout de la possibilité de spécifier un Amazon Linux personnalisé AMI lorsque vous créez un cluster. Pour plus d'informations, consultez la section [Utilisation d'une personnalisation AMI](#).

Modifications, améliorations et problèmes résolus

- HBase
 - Ajout de la possibilité de configurer des clusters HBase en lecture et réplication. Consultez [Utilisation d'un cluster de réplica en lecture](#).
 - Correctifs de nombreux bogues et améliorations
- Presto – ajout de la possibilité de configurer `node.properties`.
- YARN- possibilité ajoutée de configurer `container-log4j.properties`
- Sqoop - rétroporté [SQOOP-2880](#), qui introduit un argument vous permettant de définir le répertoire temporaire Sqoop.

Version 5.6.0

Les notes de mise à jour suivantes incluent des informations relatives à la EMR version 5.6.0 d'Amazon. Les modifications concernent la version EMR 5.5.0 d'Amazon.

Date de parution : 5 juin 2017

Mises à niveau

- Flink 1.2.1
- HBase1.3.1
- Mahout 0.13.0. Il s'agit de la première version de Mahout compatible avec Spark 2.x dans Amazon EMR version 5.0 et versions ultérieures.
- Spark 2.1.1

Modifications, améliorations et problèmes résolus

- Presto
 - Ajout de la possibilité d'SSLTLSactiver/sécuriser la communication entre les nœuds Presto en activant le chiffrement en transit à l'aide d'une configuration de sécurité. Pour plus d'informations, consultez [Chiffrement des données en transit](#).
 - Rétroportage de [Presto 7661](#), ce qui ajoute l'option VERBOSE à l'instruction EXPLAIN ANALYZE pour transmettre des statistiques de bas niveau plus détaillées sur un plan de la requête.

Version 5.5.3

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.5.3 d'Amazon. Les modifications ont été apportées à la version 5.5.2.

Date de parution initiale : 29 août 2018

Modifications, améliorations et problèmes résolus

- Cette version corrige une vulnérabilité de sécurité potentielle.

Version 5.5.2

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 5.5.2 d'Amazon. Les modifications ont été apportées à la version 5.5.1.

Date de parution initiale : 29 mars 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités potentielles.

Version 5.5.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.5.1 d'Amazon. Les modifications concernent la version EMR 5.5.0 d'Amazon.

Date de parution initiale : 22 janvier 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités associées à l'exécution spéculative (CVE-2017-5715, -2017-5753 et CVE -2017-5754). CVE Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Version 5.5.0

Les notes de mise à jour suivantes incluent des informations relatives à la version EMR 5.5.0 d'Amazon. Les modifications concernent la version EMR 5.4.0 d'Amazon.

Date de parution : 26 avril 2017

Mises à niveau

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1

- ZooKeeper 3.4.10

Modifications, améliorations et problèmes résolus

- Spark
 - Correctif de Spark rétroporté ([SPARK-20115](#)) permettant [DAGScheduler de recalculer tous les blocs de shuffle perdus lorsque le service de shuffle externe n'est pas disponible](#) dans la version 2.1.0 de Spark, incluse dans cette version.
- Flink
 - Flink s'appuie désormais sur Scala 2.11. Si vous utilisez Scala API et les bibliothèques, nous vous recommandons d'utiliser Scala 2.11 dans vos projets.
 - Résolution d'un problème où les valeurs par défaut HADOOP_CONF_DIR et YARN_CONF_DIR n'étaient pas définies correctement, si bien que `start-scala-shell.sh` ne fonctionnait pas. Ajout également de la possibilité de définir ces valeurs à l'aide de `env.hadoop.conf.dir` et `env.yarn.conf.dir` dans `/etc/flink/conf/flink-conf.yaml` ou la classification de configuration `flink-conf`.
 - Introduction d'une nouvelle commande EMR spécifique, `flink-scala-shell` en tant que wrapper pour `start-scala-shell.sh`. Nous vous recommandons d'utiliser cette commande au lieu de `start-scala-shell`. La nouvelle commande simplifie l'exécution. Par exemple, `flink-scala-shell -n 2` démarre un shell Flink Scala avec un parallélisme de 2 tâches.
 - Introduction d'une nouvelle commande EMR spécifique, `flink-yarn-session` en tant que wrapper pour `yarn-session.sh`. Nous vous recommandons d'utiliser cette commande au lieu de `yarn-session`. La nouvelle commande simplifie l'exécution. Par exemple, `flink-yarn-session -d -n 2` démarre une session Flink de longue durée à l'état détaché avec deux gestionnaires de tâches.
 - Le [httpclient commons adressé \(FLINK-6125\) n'est plus ombré dans](#) Flink 1.2.
- Presto
 - Ajout de la prise en charge de LDAP l'authentification. L'utilisation LDAP avec Presto sur Amazon EMR nécessite que vous autorisiez l'HTTPS accès au coordinateur Presto (`http-server.https.enabled=true` in `config.properties`). Pour plus de détails sur la configuration, consultez [LDAP l'authentification](#) dans la documentation Presto.
 - Ajout de la prise en charge de `SHOW GRANTS`.
- Amazon EMR Base Linux AMI

- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2017.03. Pour plus d'informations, consultez les notes de [mise à jour d'Amazon Linux AMI 2017.03](#).
- Python 2.6 a été supprimé de l'image Linux EMR de base d'Amazon. Python 2.7 et 3.4 sont installés par défaut. Vous pouvez installer Python 2.6 manuellement si nécessaire.

Version 5.4.0

Les notes de mise à jour suivantes incluent des informations relatives à la version EMR 5.4.0 d'Amazon. Les modifications concernent la version EMR 5.3.0 d'Amazon.

Date de parution : 8 mars 2017

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers Flink 1.2.0
- Mise à niveau vers HBase 1.3.0
- Mise à niveau vers Phoenix 4.9.0

Note

Si vous effectuez une mise à niveau d'une version antérieure d'Amazon EMR vers la EMR version 5.4.0 ou ultérieure d'Amazon et que vous utilisez l'indexation secondaire, mettez à niveau les index locaux comme décrit dans la documentation d'[Apache Phoenix](#). Amazon EMR supprime les configurations requises de la `hbase-site classification`, mais les index doivent être reemplis. La mise à niveau en ligne et hors ligne des index est prise en charge. Les mises à niveau sont par défaut en ligne, ce qui signifie que les index sont remplis de nouveau lors de l'initialisation depuis les clients Phoenix en version 4.8.0 ou supérieure. Pour spécifier des mises à niveau hors ligne, définissez la `phoenix.client.localIndexUpgrade` configuration sur `false` dans la `phoenix-site classification`, puis SSH sur le nœud principal à exécuter `psql [zookeeper] -1`.

- Mise à niveau vers Presto 0.166
- Mise à niveau vers Zeppelin 0.7.0

Modifications et améliorations

Les modifications suivantes ont été apportées aux EMR versions d'Amazon pour le label emr-5.4.0 :

- Ajout de la prise en charge des instances r4. Consultez la section [Types d'EC2instances Amazon](#).

Version 5.3.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.3.1 d'Amazon. Les modifications concernent la version EMR 5.3.0 d'Amazon.

Date de version : 7 février 2017

Des modifications mineures ont été apportées au rétroportage des correctifs Zeppelin et à la mise à jour de la valeur par défaut pour AmazonAMI. EMR

Version 5.3.0

Les notes de mise à jour suivantes incluent des informations relatives à la version EMR 5.3.0 d'Amazon. Les modifications concernent la version EMR 5.2.1 d'Amazon.

Date de parution : 26 janvier 2017

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers Hive 2.1.1
- Mise à niveau vers Hue 3.11.0
- Mise à niveau vers Spark 2.1.0
- Mise à niveau vers Oozie 4.3.0
- Mise à niveau vers Flink 1.1.4

Modifications et améliorations

Les modifications suivantes ont été apportées aux EMR versions d'Amazon pour le label de version emr-5.3.0 :

- Ajout à Hue d'un correctif qui vous permet d'utiliser le paramètre `interpreters_shown_on_wheel` pour configurer les interpréteurs à afficher en premier sur la roue de sélection Notebook quel que soit leur ordre dans le fichier `hue.ini`.
- Ajout de la classification de configuration, `hive-parquet-logging` que vous pouvez utiliser pour configurer des valeurs dans le fichier `parquet-logging.properties` de Hive.

Version 5.2.2

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.2.2 d'Amazon. Les modifications concernent la version EMR 5.2.1 d'Amazon.

Date de parution : 2 mai 2017

Problèmes connus résolus depuis les versions précédentes

- Rétroporté [SPARK-194459](#), qui résout un problème en cas d'échec de la lecture d'une ORC table contenant des colonnes `char`/`varchar`.

Version 5.2.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.2.1 d'Amazon. Les modifications concernent la version EMR 5.2.0 d'Amazon.

Date de parution : 29 décembre 2016

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers Presto 0.157.1. Pour plus d'informations, consultez [Notes de mise à jour Presto](#) dans la documentation Presto.
- Mise à niveau vers Zookeeper 3.4.9. Pour plus d'informations, consultez les [notes ZooKeeper de publication](#) dans la ZooKeeper documentation d'Apache.

Modifications et améliorations

Les modifications suivantes ont été apportées aux EMR versions d'Amazon pour le label de version `emr-5.2.1` :

- Ajout de la prise en charge du type d'EC2instance Amazon m4.16xlarge dans les EMR versions 4.8.3 et ultérieures d'Amazon, à l'exception des versions 5.0.0, 5.0.3 et 5.2.0.
- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.09. Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.
- L'emplacement de Flink et les chemins de YARN configuration sont désormais définis par défaut, /etc/default/flink ce qui signifie qu'il n'est pas nécessaire de définir les variables d'environnement FLINK_CONF_DIR et HADOOP_CONF_DIR lorsque vous exécutez les scripts du `yarn-session.sh` pilote `flink` or pour lancer les tâches Flink.
- Ajout du support pour les FlinkKinesisConsumer cours.

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème dans Hadoop où le ReplicationMonitor thread pouvait rester bloqué pendant longtemps en raison d'une course entre la réplication et la suppression du même fichier dans un grand cluster.
- Correction d'un problème en raison duquel ControlledJob # toString échouait avec une exception de pointeur nul (NPE) lorsque le statut de la tâche n'était pas correctement mis à jour.

Version 5.2.0

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.2.0 d'Amazon. Les modifications concernent la version EMR 5.1.0 d'Amazon.

Date de parution : 21 novembre 2016

Modifications et améliorations

Les modifications et les améliorations suivantes sont disponibles dans cette version :

- Ajout du mode de stockage Amazon S3 pourHBase.
- Vous permet de spécifier un emplacement Amazon S3 pour le HBase rootdir. Pour plus d'informations, consultez [HBaseAmazon S3](#).

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers Spark 2.0.2

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème lié à la limitation de /mnt à 2 To pour les types d'instance EBS réservés uniquement.
- Correction d'un problème avec les journaux instance-controller et logpusher édités dans leurs fichiers .out correspondants au lieu de leurs fichiers .log configurés log4j, qui permutent toutes les heures. Les fichiers .out n'effectuent pas de permutation, cela remplirait finalement la partition /emr. Ce problème concerne uniquement les types d'instances matérielles de machine virtuelle (HVM).

Version 5.1.0

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.1.0 d'Amazon. Les modifications concernent la version EMR 5.0.0 d'Amazon.

Date de parution : 3 novembre 2016

Modifications et améliorations

Les modifications et les améliorations suivantes sont disponibles dans cette version :

- Support ajouté pour Flink 1.1.3.
- Presto a été ajouté comme une option dans la section ordinateur portable de Hue.

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers la version HBase 1.2.3
- Mise à niveau vers Zeppelin 0.6.2

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème lié aux requêtes Tez sur Amazon S3 avec des ORC fichiers qui ne fonctionnaient pas aussi bien que dans les versions antérieures d'Amazon EMR 4.x.

Version 5.0.3

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 5.0.3 d'Amazon. Les modifications concernent la version EMR 5.0.0 d'Amazon.

Date de parution : 24 octobre 2016

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers Hadoop 2.7.3
- Mise à niveau vers Presto 0.152.3, qui comprend la prise en charge de l'interface Web Presto. Utilisez le port 8889 du coordinateur Presto pour accéder à l'interface Web Presto. Pour plus d'informations sur l'interface Web Presto, consultez [Interface Web](#) dans la documentation Presto.
- Mise à niveau vers Spark 2.0.1
- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.09. Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Version 5.0.0

Date de parution : 27 juillet 2016

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers Hive 2.1
- Mise à niveau vers Presto 0.150
- Mise à niveau vers Spark 2.0
- Mise à niveau vers Hue 3.10.0
- Mise à niveau vers Pig 0.16.0
- Mise à niveau vers Tez 0.8.4
- Mise à niveau vers Zeppelin 0.6.1

Modifications et améliorations

Les modifications suivantes ont été apportées aux EMR versions Amazon pour le label de version emr-5.0.0 ou supérieur :

- Amazon EMR prend en charge les dernières versions open source de Hive (version 2.1) et Pig (version 0.16.0). Si vous avez déjà utilisé Hive ou Pig sur AmazonEMR, cela peut affecter certains cas d'utilisation. Pour en savoir plus, consultez [Hive](#) et [Pig](#).
- Tez est désormais le moteur d'exécution par défaut de Hive et Pig. Pour changer cela, vous devez modifier les valeurs appropriées dans les classifications de configuration `hive-site` et `pig-properties`, respectivement.
- Une fonction améliorée de débogage des étapes a été ajoutée, ce qui vous permet de voir la cause première des échecs des étapes si le service peut déterminer la cause. Pour plus d'informations, consultez la section [Enhanced Step Debugging](#) dans le Amazon EMR Management Guide.
- Les applications qui se terminaient par « -Sandbox » n'ont plus ce suffixe. Cela peut interrompre l'automatisation de certaines tâches, par exemple, si vous utilisez des scripts pour lancer des clusters avec ces applications. Le tableau suivant montre les noms des applications dans Amazon EMR 4.7.2 par rapport à Amazon EMR 5.0.0.

Modification des noms d'application

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Bac à sable	ZooKeeper

- Spark est maintenant compilé pour Scala 2.11.
- Java 8 est désormais la version par défaut JVM. Toutes les applications s'exécutent avec le runtime Java 8. Aucune modification n'est apportée bytecode cible des applications. La plupart des applications continuent de cibler Java 7.
- Zeppelin inclut désormais des fonctions d'authentification. Pour en savoir plus, consultez [Zeppelin](#).

- Ajout de la prise en charge des configurations de sécurité, qui vous permettent de créer et d'appliquer des options de chiffrement plus facilement. Pour en savoir plus, consultez [Chiffrement des données](#).

Version 4.9.5

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 4.9.5 d'Amazon. Les modifications ont été apportées à la version 4.9.4.

Date de parution initiale : 29 août 2018

Modifications, améliorations et problèmes résolus

- HBase
 - Cette version corrige une vulnérabilité de sécurité potentielle.

Version 4.9.4

Les notes de mise à jour suivantes contiennent des informations relatives à la EMR version 4.9.4 d'Amazon. Les modifications ont été apportées à la version 4.9.3.

Date de parution initiale : 29 mars 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités potentielles.

Version 4.9.3

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.9.3 d'Amazon. Les modifications concernent la version EMR 4.9.2 d'Amazon.

Date de parution initiale : 22 janvier 2018

Modifications, améliorations et problèmes résolus

- Mise à jour du noyau Amazon Linux de defaultAmazon Linux AMI pour Amazon EMR afin de corriger les vulnérabilités associées à l'exécution spéculative (CVE-2017-5715, -2017-5753 et CVE

-2017-5754). CVE Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Version 4.9.2

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.9.2 d'Amazon. Les modifications concernent la version EMR 4.9.1 d'Amazon.

Date de parution : 13 juillet 2017

Des modifications mineures, des correctifs de bogues et des améliorations ont été apportées à cette version.

Version 4.9.1

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.9.1 d'Amazon. Les modifications concernent la version EMR 4.8.4 d'Amazon.

Date de parution : 10 avril 2017

Problèmes connus résolus depuis les versions précédentes

- [Rétroportages de HIVE-9976 et -10106 HIVE](#)
- Correction d'un problème à cause YARN duquel un grand nombre de nœuds (supérieur à 2 000) et de conteneurs (supérieur à 5 000) provoquait une erreur de mémoire insuffisante, par exemple :
`"Exception in thread 'main' java.lang.OutOfMemoryError"`

Modifications et améliorations

Les modifications suivantes ont été apportées aux EMR versions d'Amazon pour le label de version emr-4.9.1 :

- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2017.03. Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- Python 2.6 a été supprimé de l'image Linux EMR de base d'Amazon. Vous pouvez installer Python 2.6 manuellement si nécessaire.

Version 4.8.4

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.8.4 d'Amazon. Les modifications concernent la version EMR 4.8.3 d'Amazon.

Date de parution : 7 février 2017

Des modifications mineures, des correctifs de bogues et des améliorations ont été apportées à cette version.

Version 4.8.3

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.8.3 d'Amazon. Les modifications concernent la version EMR 4.8.2 d'Amazon.

Date de parution : 29 décembre 2016

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers Presto 0.157.1. Pour plus d'informations, consultez [Notes de mise à jour Presto](#) dans la documentation Presto.
- Mise à niveau vers Spark 1.6.3. Pour plus d'informations, consultez les [Notes de mise à jour Spark](#) dans la documentation Apache Spark.
- Mise à niveau vers la version ZooKeeper 3.4.9. Pour plus d'informations, consultez les [notes ZooKeeper de publication](#) dans la ZooKeeper documentation d'Apache.

Modifications et améliorations

Les modifications suivantes ont été apportées aux EMR versions d'Amazon pour le label de version emr-4.8.3 :

- Ajout de la prise en charge du type d'EC2instance Amazon m4.16xlarge dans les EMR versions 4.8.3 et ultérieures d'Amazon, à l'exception des versions 5.0.0, 5.0.3 et 5.2.0.
- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.09. Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème dans Hadoop où le ReplicationMonitor thread pouvait rester bloqué pendant longtemps en raison d'une course entre la réplication et la suppression du même fichier dans un grand cluster.
- Correction d'un problème en raison duquel ControlledJob # toString échouait avec une exception de pointeur nul (NPE) lorsque le statut de la tâche n'était pas correctement mis à jour.

Version 4.8.2

Les notes de mise à jour suivantes contiennent des informations relatives à la version EMR 4.8.2 d'Amazon. Les modifications concernent la version EMR 4.8.0 d'Amazon.

Date de parution : 24 octobre 2016

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers Hadoop 2.7.3
- Mise à niveau vers Presto 0.152.3, qui comprend la prise en charge de l'interface Web Presto. Utilisez le port 8889 du coordinateur Presto pour accéder à l'interface Web Presto. Pour plus d'informations sur l'interface Web Presto, consultez [Interface Web](#) dans la documentation Presto.
- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.09. Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Version 4.8.0

Date de parution : 7 septembre 2016

Mises à niveau

Les mises à niveau suivantes sont disponibles dans cette version :

- Mise à niveau vers la version HBase 1.2.2
- Mise à niveau vers Presto-Sandbox 0.151

- Mise à niveau vers Tez 0.8.4
- Mise à niveau vers Zeppelin-Sandbox 0.6.1

Modifications et améliorations

Les modifications suivantes ont été apportées aux EMR versions d'Amazon pour le label de version emr-4.8.0 :

- Correction d'un problème en raison YARN duquel ils ApplicationMaster tentaient de nettoyer des conteneurs qui n'existent plus parce que leurs instances ont été résiliées.
- Corrigé les actions hive-server2 URL pour Hive2 dans les exemples d'Oozie.
- Ajout de la prise en charge d'autres catalogues Presto.
- [Correctifs rétroportés : HIVE-8948, HIVE -12679, -13405, -3116, HIVE -12689 PHOENIX HADOOP](#)
- Ajout de la prise en charge des configurations de sécurité, qui vous permettent de créer et d'appliquer des options de chiffrement plus facilement. Pour en savoir plus, consultez [Chiffrement des données](#).

Version 4.7.2

Les notes de mise à jour suivantes contiennent des informations relatives à Amazon EMR 4.7.2.

Date de parution : 15 juillet 2016

Fonctionnalités

Les fonctions suivantes sont disponibles dans cette version :

- Mise à niveau vers Mahout 0.12.2
- Mise à niveau vers Presto 0.148
- Mise à niveau vers Spark 1.6.2
- Vous pouvez désormais créer un AWSCredentialsProvider à utiliser en EMRFS utilisant a URI comme paramètre. Pour plus d'informations, consultez la section [Créer un AWSCredentialsProvider formulaire EMRFS](#).
- EMRFS permet désormais aux utilisateurs de configurer un point de terminaison DynamoDB personnalisé pour leurs métadonnées Consistent View à l'aide `fs.s3.consistent.dynamodb.endpoint` de la propriété `in.emrfs-site.xml`

- Ajout d'un script dans `/usr/bin` appelé `spark-example`, qui encapsule `/usr/lib/spark/spark/bin/run-example` pour vous permettre d'exécuter des exemples directement. Par exemple, pour exécuter l' SparkPi exemple fourni avec la distribution Spark, vous pouvez l'exécuter `spark-example SparkPi 100` à partir de la ligne de commande ou en utilisant `command-runner.jar` comme étape leAPI.

Problèmes connus résolus depuis les versions précédentes

- Résolution d'un problème où `spark-assembly.jar` pour Oozie n'était pas à l'emplacement approprié quand Spark était également installé, ce qui provoquait un échec du lancement d'applications Spark avec Oozie.
- Correction d'un problème lié à la journalisation dans les conteneurs basée sur Spark Log4J. YARN

Version 4.7.1

Date de parution : 10 juin 2016

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème qui prolongeait le temps de démarrage des clusters lancés dans un environnement VPC doté de sous-réseaux privés. Le bogue ne concernait que les clusters lancés avec la version EMR 4.7.0 d'Amazon.
- Correction d'un problème qui entraînait une gestion incorrecte de la liste des fichiers dans Amazon EMR pour les clusters lancés avec la EMR version 4.7.0 d'Amazon.

Version 4.7.0

Important

Amazon EMR 4.7.0 est obsolète. Utilisez Amazon EMR 4.7.1 ou version ultérieure à la place.

Date de parution : 2 juin 2016

Fonctionnalités

Les fonctions suivantes sont disponibles dans cette version :

- Ajout d'Apache Phoenix 4.7.0
- Ajout d'Apache Tez 0.8.3
- Mise à niveau vers la version HBase 1.2.1
- Mise à niveau vers Mahout 0.12.0
- Mise à niveau vers Presto 0.147
- Mise à niveau de la AWS SDK for Java version 1.10.75
- L'indicateur final a été supprimé de la propriété `mapreduce.cluster.local.dir` dans `mapred-site.xml` pour permettre aux utilisateurs d'exécuter Pig en mode local.

JDBCPilotes Amazon Redshift disponibles sur le cluster

Les JDBC pilotes Amazon Redshift sont désormais inclus sur `/usr/share/aws/redshift/jdbc`. `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` est le pilote Amazon JDBC Redshift compatible 4.1 `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` et le pilote Amazon Redshift JDBC compatible 4.0. Pour plus d'informations, consultez [Configurer une JDBC connexion](#) dans le guide de gestion Amazon Redshift.

Java 8

À l'exception de Presto, Open JDK 1.7 est la version par défaut JDK utilisée pour toutes les applications. Cependant, Open JDK 1.7 et 1.8 sont tous deux installés. Pour plus d'informations sur la configuration de `JAVA_HOME` pour les applications, consultez [Configuration d'applications pour utiliser Java 8](#).

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème de noyau qui affectait de manière significative les performances des EBS volumes optimisés pour le débit HDD (st1) pour Amazon EMR dans `emr-4.6.0`.
- Correction d'un problème en raison duquel un cluster échouait si une zone de HDFS chiffrement était spécifiée sans choisir Hadoop comme application.
- La politique d'HDFSécriture par défaut est passée de `RoundRobin` à `AvailableSpaceVolumeChoosingPolicy`. Certains volumes n'ont pas été correctement utilisés avec la `RoundRobin` configuration, ce qui a entraîné la défaillance des nœuds principaux et un manque de fiabilité HDFS.
- Correction d'un problème lié au `EMRFSCCLI`, qui provoquait une exception lors de la création de la table de métadonnées DynamoDB par défaut pour des vues cohérentes.

- Correction d'un problème de blocage EMRFS qui se produisait potentiellement lors d'opérations de renommage et de copie en plusieurs parties.
- Correction d'un problème en EMRFS raison duquel la CopyPart taille par défaut était de 5 Mo. La valeur par défaut est maintenant correctement définie sur 128 Mo.
- Résolution d'un problème lié à la configuration upstart dans Zeppelin qui pouvait vous empêcher potentiellement d'arrêter le service.
- Correction d'un problème lié à Spark et Zeppelin, qui vous empêchait d'utiliser le s3a:// URI schéma car il n'y avait pas correctement chargé dans leur chemin de classe respectif.
- Rétroporté [HUE-2484](#).
- J'ai rétroporté un [commit](#) depuis Hue 3.9.0 (JIRAil n'existe pas) pour corriger un problème avec l'exemple de HBase navigateur.
- Rétroporté [HIVE-9073](#).

Version 4.6.0

Date de parution : 21 avril 2016

Fonctionnalités

Les fonctions suivantes sont disponibles dans cette version :

- Ajouté HBase 1.2.0
- Ajout de Zookeeper-Sandbox 3.4.8
- Mise à niveau vers Presto-Sandbox 0.143
- Les EMR versions d'Amazon sont désormais basées sur Amazon Linux 2016.03.0. Pour de plus amples informations, veuillez consulter <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.

Problème affectant les types de volumes optimisés pour le débit HDD (st1) EBS

Un problème lié aux versions 4.2 et supérieures du noyau Linux affecte de manière significative les performances des EBS volumes optimisés pour le débit HDD (st1) pour. EMR Cette version (emr-4.6.0) utilise la version de noyau 4.4.5 et est donc concernée. Par conséquent, nous vous

recommandons de ne pas utiliser `emr-4.6.0` si vous souhaitez utiliser des volumes `st1`. EBS Vous pouvez utiliser `emr-4.5.0` ou des EMR versions antérieures d'Amazon avec `st1` sans impact. En outre, nous fournirons le correctif avec les versions futures.

Valeurs par défaut de Python

Python 3.4 est maintenant installé par défaut, mais Python 2.7 reste la valeur système par défaut. Vous pouvez configurer Python 3.4 comme valeur par défaut du système à l'aide d'une action bootstrap ; vous pouvez utiliser la configuration API pour définir `PYSPARK_PYTHON` export to `/usr/bin/python3.4` dans la `spark-env` classification afin d'affecter la version de Python utilisée par PySpark.

Java 8

À l'exception de Presto, Open JDK 1.7 est la version par défaut JDK utilisée pour toutes les applications. Cependant, Open JDK 1.7 et 1.8 sont tous deux installés. Pour plus d'informations sur la configuration de `JAVA_HOME` pour les applications, consultez [Configuration d'applications pour utiliser Java 8](#).

Problèmes connus résolus depuis les versions précédentes

- Résolution d'un problème où le provisionnement d'application échouait parfois de manière aléatoire en raison d'un mot de passe généré.
- Auparavant, `mysqld` était installé sur tous les nœuds. Maintenant, il est uniquement installé sur l'instance principale et seulement si l'application choisie inclut `mysql-server` comme composant. Actuellement, les applications suivantes incluent le `mysql-server` composant : HiveHCatalog, Hue, Presto-Sandbox et Sqoop-Sandbox.
- La valeur par défaut est passée `yarn.scheduler.maximum-allocation-vcores` à 80 par rapport à 32, ce qui corrige un problème introduit dans `emr-4.4.0` qui se produit principalement avec Spark lors de l'utilisation de `maximizeResourceAllocationOption` dans un cluster dont le type d'instance principal est l'un des rares types d'instance dont le nombre de YARN cœurs est supérieur à 32, à savoir `c4.8xlarge`, `d2.8xlarge`, `d2.8xlarge` ou `m4.10xlarge` ont été concernés par ce problème.
- `s3-dist-cp` est désormais utilisé pour toutes les nominations EMRFS Amazon S3 et n'est plus transféré vers un répertoire temporaire. HDFS
- Résolution d'un problème lié au traitement des exceptions pour le chiffrement côté client du chargement partitionné.

- Ajout d'une option pour permettre aux utilisateurs de modifier la classe de stockage Amazon S3. Par défaut, ce paramètre est STANDARD. La configuration de classification `emrfs-site` est `fs.s3.storageClass`, et les valeurs possibles sont STANDARD, STANDARD_IA et REDUCED_REDUNDANCY. Pour plus d'informations sur les classes de stockage, consultez la section [Classes de stockage](#) dans le Guide de l'utilisateur Amazon Simple Storage Service.

Version 4.5.0

Date de parution : 4 avril 2016

Fonctionnalités

Les fonctions suivantes sont disponibles dans cette version :

- Mise à niveau vers Spark 1.6.1
- Mise à niveau vers Hadoop 2.7.2
- Mise à niveau vers Presto 0.140
- Ajout de la AWS KMS prise en charge du chiffrement côté serveur Amazon S3.

Problèmes connus résolus depuis les versions précédentes

- Problème résolu : les serveurs My SQL et Apache ne démarraient pas après le redémarrage d'un nœud.
- Correction d'un problème en raison duquel les tables non partitionnées stockées dans Amazon S3 IMPORT ne fonctionnaient pas correctement
- Résolution d'un problème lié à Presto où le répertoire intermédiaire devait être `/mnt/tmp` au lieu de `/tmp` lors de l'écriture dans des tables Hive.

Version 4.4.0

Date de parution : 14 mars 2016

Fonctionnalités

Les fonctions suivantes sont disponibles dans cette version :

- Ajouté HCatalog 1.0.0

- Ajout de Sqoop-Sandbox 1.4.6
- Mise à niveau vers Presto 0.136
- Mise à niveau vers Zeppelin 0.5.6
- Mise à niveau vers Mahout 0.11.1
- `dynamicResourceAllocation` activé par défaut.
- Ajout d'un tableau de toutes les classifications de configuration pour la version. Pour en savoir plus, consultez le tableau Classifications des configurations dans [Configuration des applications](#).

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème à cause duquel le `maximizeResourceAllocation` paramètre ne réservait pas suffisamment de mémoire aux YARN ApplicationMaster démons.
- Correction d'un problème rencontré avec une personnalisation DNS. Si des entrées dans `resolve.conf` précèdent les entrées personnalisées fournies, les entrées personnalisées ne peuvent pas être résolues. Ce comportement a été affecté par les clusters dans VPC lesquels le serveur de VPC noms par défaut est inséré en tant qu'entrée supérieure `resolve.conf`.
- Résolution d'un problème où la version Python par défaut passait à la version 2.7 et boto n'était pas installé pour cette version.
- Correction d'un problème à cause duquel les YARN conteneurs et les applications Spark généraient un fichier de base de données Ganglia Round Robin (rrd) unique, ce qui entraînait le remplissage du premier disque attaché à l'instance. En raison de ce correctif, les métriques au niveau du YARN conteneur ont été désactivées et les métriques au niveau de l'application Spark ont été désactivées.
- Résolution d'un problème dans le transmetteur de journaux où celui-ci supprimait tous les dossiers de journal vides. Cela a eu pour effet que la Hive n'CLl a pas pu se connecter car Log Pusher supprimait le `user` dossier vide situé en dessous. `/var/log/hive`
- Résolution d'un problème affectant les importations Hive qui avait une incidence sur le partitionnement et entraînait une erreur lors de l'importation.
- Correction d'un problème à cause duquel EMRFS `s3-dist-cp` ne gérait pas correctement les noms de compartiments contenant des points.
- Modification d'un comportement de EMRFS telle sorte que dans les compartiments compatibles avec les versions, le fichier de `_$folder$` marqueurs n'est pas créé en permanence, ce qui peut contribuer à améliorer les performances des compartiments compatibles avec les versions.

- Le comportement a été modifié de EMRFS telle sorte qu'il n'utilise pas de fichiers d'instructions, sauf dans les cas où le chiffrement côté client est activé. Si vous souhaitez supprimer des fichiers d'instruction tout en utilisant le chiffrement côté client, vous pouvez définir la propriété `emrfs-site.xml fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled` sur `true`.
- Modification de l'agrégation des YARN journaux pour conserver les journaux à la destination de l'agrégation pendant deux jours. La destination par défaut est le HDFS stockage de votre cluster. Si vous souhaitez modifier cette durée, remplacez la valeur `yarn.log-aggregation.retain-seconds` à l'aide de la classification de configuration `yarn-site` lorsque vous créez votre cluster. Comme toujours, vous pouvez enregistrer vos journaux d'applications dans Amazon S3 à l'aide du paramètre `log-uri` lorsque vous créez votre cluster.

Correctifs appliqués

Les correctifs suivants de projets open source ont été inclus dans cette version :

- [HIVE-9655](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

Version 4.3.0

Date de parution : 19 janvier 2016

Fonctionnalités

Les fonctions suivantes sont disponibles dans cette version :

- Mise à niveau vers Hadoop 2.7.1
- Mise à niveau vers Spark 1.6.0
- Mise à niveau de Ganglia vers la version 3.7.2
- Mise à niveau vers de Presto vers la version 0.130

Amazon EMR a apporté quelques modifications au `spark.dynamicAllocation.enabled` moment où il est défini sur `true` ; il est défini sur `false` par défaut. Lorsque ce paramètre est défini sur `true`, cela a un effet sur les valeurs par défaut définies par le paramètre `maximizeResourceAllocation`.

- Si `spark.dynamicAllocation.enabled` est défini sur `true`, `spark.executor.instances` n'est pas défini par `maximizeResourceAllocation`.
- Le paramètre `spark.driver.memory` est désormais configuré en fonction des types d'instances du cluster d'une manière similaire à la façon dont `spark.executors.memory` est défini. Cependant, étant donné que l'application du pilote Spark peut s'exécuter sur l'instance principale ou sur l'une des instances principales (par exemple, en mode YARN client et en mode cluster, respectivement), le `spark.driver.memory` paramètre est défini en fonction du type d'instance du type d'instance le plus petit entre ces deux groupes d'instances.
- Le `spark.default.parallelism` paramètre est désormais défini sur le double du nombre de CPU cœurs disponibles pour les YARN conteneurs. Dans les versions précédentes, c'était la moitié de cette valeur.
- Les calculs de la surcharge de mémoire réservée aux YARN processus Spark ont été ajustés pour être plus précis, ce qui a entraîné une légère augmentation de la quantité totale de mémoire disponible pour Spark (c'est-à-dire, `spark.executor.memory`).

Problèmes connus résolus depuis les versions précédentes

- YARN l'agrégation des journaux est désormais activée par défaut.
- Correction d'un problème en raison duquel les journaux n'étaient pas transférés vers le compartiment de journaux Amazon S3 d'un cluster lorsque l'agrégation des YARN journaux était activée.
- YARN les tailles de conteneur ont désormais un nouveau minimum de 32 pour tous les types de nœuds.
- Résolution d'un problème lié à Ganglia qui entraînait des E/S de disque excessives sur le nœud principal dans les clusters de grande taille.
- Résolution d'un problème qui empêchait les journaux des applications d'être transmis à Amazon S3 lorsqu'un cluster s'arrêtait.
- Correction d'un problème EMRFS CLI qui provoquait l'échec de certaines commandes.
- Correction d'un problème lié à Zeppelin qui empêchait le chargement des dépendances dans le sous-jacent. `SparkContext`
- Résolution d'un problème provoqué par une tentative de redimensionnement pour ajouter des instances.
- Correction d'un problème dans Hive à cause duquel `CREATE TABLE AS SELECT` envoyait trop d'appels de liste à Amazon S3.

- Résolution d'un problème où les grands clusters n'étaient pas provisionnés correctement lorsque Hue, Oozie et Ganglia étaient installés.
- Résolution d'un problème dans s3-dist-cp où un code de sortie zéro était renvoyé même en cas d'échec avec une erreur.

Correctifs appliqués

Les correctifs suivants de projets open source ont été inclus dans cette version :

- [OOZIE-2402](#)
- [HIVE-12502](#)
- [HIVE-10631](#)
- [HIVE-12213](#)
- [HIVE-10559](#)
- [HIVE-12715](#)
- [HIVE-10685](#)

Version 4.2.0

Date de parution : 18 novembre 2015

Fonctionnalités

Les fonctions suivantes sont disponibles dans cette version :

- Ajout de la prise en charge de Ganglia
- Mise à niveau vers Spark 1.5.2
- Mise à niveau vers Presto 0.125
- Mise à niveau d'Oozie vers la version 4.2.0
- Mise à niveau de Zeppelin vers la version 0.5.5
- Mise à niveau de la AWS SDK for Java version 1.10.27

Problèmes connus résolus depuis les versions précédentes

- Correction d'un problème en raison EMRFS CLI duquel le nom de la table de métadonnées par défaut n'était pas utilisé.
- Correction d'un problème rencontré lors de l'utilisation de tables ORC sauvegardées par -back dans Amazon S3.
- Résolution d'un problème rencontré avec une incompatibilité de version Python dans la configuration de Spark.
- Correction d'un problème lorsque l'état d'un YARN nœud n'était pas signalé en raison de DNS problèmes liés aux clusters d'unVPC.
- Correction d'un problème rencontré lors de la YARN mise hors service de nœuds, qui entraînait le blocage d'applications ou l'impossibilité de planifier de nouvelles applications.
- Correction d'un problème rencontré lorsque les clusters se terminaient avec le statut TIMED _ OUT _STARTING.
- Correction d'un problème rencontré lors de l'inclusion de la dépendance EMRFS Scala dans d'autres versions. La dépendance Scala a été supprimée.

Configuration des applications

Pour remplacer les configurations par défaut d'une application, vous pouvez fournir un objet de configuration. Vous pouvez soit utiliser une syntaxe abrégée pour fournir la configuration, soit référencer l'objet de configuration dans un JSON fichier. Les objets de configuration sont composés d'une classification, de propriétés et de configurations imbriquées en option. Les propriétés correspondent aux paramètres de l'application que vous souhaitez modifier. Vous pouvez définir plusieurs classifications pour plusieurs applications dans un seul JSON objet.

Warning

Les API opérations Amazon EMR Describe et List émettent des paramètres personnalisés et configurables, qui sont utilisés dans le cadre des flux de EMR travail Amazon, en texte clair. Pour fournir des informations sensibles, telles que des mots de passe, dans ces paramètres, consultez [Stocker les données de configuration sensibles dans AWS Secrets Manager](#).

Les classifications de configuration disponibles varient selon la EMR version d'Amazon. Pour obtenir la liste des classifications de configuration prises en charge dans une version de version particulière, reportez-vous à la page de cette version sous [À propos d'Amazon EMR Releases](#).

Voici un exemple de JSON fichier pour une liste de configurations.

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.security.groups.cache.secs": "250"
    }
  },
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.tasktracker.map.tasks.maximum": "2",
      "mapreduce.map.sort.spill.percent": "0.90",
      "mapreduce.tasktracker.reduce.tasks.maximum": "5"
    }
  }
]
```

Une classification de configuration correspond généralement à un fichier de configuration spécifique à l'application. Par exemple, la classification `hive-site` est mappée aux paramètres du fichier de configuration `hive-site.xml` pour Hive. Une exception à cette règle est l'action d'amorçage `configure-daemons`, qui n'est plus prise en charge et qui est utilisée pour définir des paramètres d'environnement tels que `--namenode-heap-size`. Les options de ce type sont intégrées dans les classifications `hadoop-env` et `yarn-env` avec leurs propres classifications d'exportation imbriquées. Si des classifications se terminent par `env`, utilisez la sous-classification d'exportation.

`s3get` est une autre exception qui est utilisée pour placer un objet

`EncryptionMaterialsProvider` client sur chaque nœud d'un cluster pour l'utiliser dans le chiffrement côté client. Une option a été ajoutée à la classification `emrfs-site` à cet effet.

Voici un exemple de la classification `hadoop-env`.

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

Voici un exemple de la classification `yarn-env`.

```
[
  {
    "Classification": "yarn-env",
    "Properties": {
```

```

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "YARN_RESOURCEMANAGER_OPTS": "-Xdebug -Xrunjdp:transport=dt_socket"
        },
        "Configurations": [
          ]
        ]
      }
    ]
  }
]

```

Les paramètres suivants n'appartiennent pas à un fichier de configuration mais sont utilisés par Amazon EMR pour éventuellement configurer plusieurs paramètres en votre nom.

Paramètres sélectionnés par Amazon EMR

Application	Classification d'étiquette de version	Propriétés valides	Utilisation
Spark	spark	maximizeResourceAllocation	Configurez des programmes d'exécution pour utiliser le maximum de ressources de chaque nœud.

Rubriques

- [Configuration des applications lorsque vous créez un cluster](#)
- [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#)
- [Stocker les données de configuration sensibles dans AWS Secrets Manager](#)
- [Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique](#)

Configuration des applications lorsque vous créez un cluster

Lorsque vous créez un cluster, vous pouvez remplacer les configurations par défaut pour les applications utilisant la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK.

Pour remplacer la configuration par défaut d'une application, vous devez spécifier des valeurs personnalisées dans une classification de configuration. Une classification de configuration correspond à un XML fichier de configuration pour une application, tel que `hive-site.xml`.

Les classifications de configuration varient selon la EMR version d'Amazon. Pour obtenir une liste des classifications de configuration disponibles dans une version spécifique, consultez la page des détails de la version. Par exemple, [Amazon EMR version 6.4.0](#).

Fournir une configuration dans la console lorsque vous créez un cluster

Pour fournir une configuration, accédez à la page Créer un cluster et développez les paramètres du logiciel. Vous pouvez ensuite entrer la configuration directement en utilisant l'une JSON ou l'autre syntaxe abrégée illustrée dans le texte masqué dans la console. Sinon, vous pouvez fournir un Amazon S3 URI pour un fichier contenant un JSON Configurations objet.

Pour fournir une configuration pour un groupe d'instances, choisissez un cluster dans votre liste de clusters, puis cliquez sur l'onglet Configurations. Dans le tableau des configurations des groupes d'instances, choisissez le groupe d'instances à modifier, puis choisissez Reconfigurer.

Fournissez une configuration à l'aide du AWS CLI lorsque vous créez un cluster

Vous pouvez fournir une configuration à `create-cluster` en fournissant un chemin d'accès à un JSON fichier stocké localement ou dans Amazon S3. L'exemple suivant suppose que vous utilisez des rôles par défaut pour Amazon EMR et que les rôles ont été créés. Si vous devez créer les rôles, exécutez d'abord `aws emr create-default-roles`.

Si votre configuration se trouve dans votre répertoire local, vous pouvez utiliser l'exemple de commande suivant.

```
aws emr create-cluster --use-default-roles --release-label emr-7.2.0 --applications
Name=Hive \
```

```
--instance-type m5.xlarge --instance-count 3 --configurations file:///./  
configurations.json
```

Si votre configuration se trouve dans un chemin Amazon S3, vous devez configurer la solution de contournement suivante avant de transmettre le chemin Amazon S3 à la commande `create-cluster`.

```
#!/bin/sh  
# Assume the ConfigurationS3Path is not public, and its present in the same AWS account  
# as the EMR cluster  
ConfigurationS3Path="s3://my-bucket/config.json"  
# Get a presigned HTTP URL for the s3Path  
ConfigurationURL=`aws s3 presign $ConfigurationS3Path --expires-in 300`  
# Fetch the presigned URL, and minify the JSON so that it spans only a single line  
Configurations=`curl $ConfigurationURL | jq -c .`  
aws emr create-cluster --use-default-roles --release-label emr-5.34.0 --instance-type  
m5.xlarge --instance-count 2 --applications Name=Hadoop Name=Spark --configurations  
$Configurations
```

Fournissez une configuration à l'aide de Java SDK lorsque vous créez un cluster

L'extrait de programme suivant montre comment fournir une configuration à l'aide de AWS SDK for Java.

```
Application hive = new Application().withName("Hive");  
  
Map<String,String> hiveProperties = new HashMap<String,String>();  
hiveProperties.put("hive.join.emit.interval","1000");  
hiveProperties.put("hive.merge.mapfiles","true");  
  
Configuration myHiveConfig = new Configuration()  
    .withClassification("hive-site")  
    .withProperties(hiveProperties);  
  
RunJobFlowRequest request = new RunJobFlowRequest()  
    .withName("Create cluster with ReleaseLabel")  
    .withReleaseLabel("emr-5.20.0")  
    .withApplications(hive)  
    .withConfigurations(myHiveConfig)  
    .withServiceRole("EMR_DefaultRole")
```

```
.withJobFlowRole("EMR_EC2_DefaultRole")
.withInstances(new JobFlowInstancesConfig()
    .withEc2KeyName("myEc2Key")
    .withInstanceCount(3)
    .withKeepJobFlowAliveWhenNoSteps(true)
    .withMasterInstanceType("m4.large")
    .withSlaveInstanceType("m4.large")
);
```

Reconfigurer un groupe d'instances dans un cluster en cours d'exécution

Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez reconfigurer les applications de cluster et spécifier des classifications de configuration supplémentaires pour chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, vous pouvez utiliser la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK.

Lorsque vous mettez à jour la configuration d'une application pour un groupe d'instances dans la nouvelle EMR console Amazon, celle-ci tente de fusionner la nouvelle configuration avec la configuration existante pour créer une nouvelle configuration active. Dans le cas inhabituel où Amazon ne parvient pas à fusionner la configuration, la console vous avertit.

Après avoir soumis une demande de reconfiguration pour un groupe d'instances, Amazon EMR attribue un numéro de version à la nouvelle spécification de configuration. Vous pouvez suivre le numéro de version d'une configuration ou l'état d'un groupe d'instances en consultant les CloudWatch événements. Pour plus d'informations, consultez la section [Surveiller CloudWatch les événements](#).

Note

Vous pouvez uniquement remplacer, et non supprimer, les configurations de cluster spécifiées lors de la création du cluster. S'il existe des différences entre la configuration existante et le fichier que vous fournissez, Amazon EMR rétablit les configurations modifiées manuellement, telles que les configurations que vous avez modifiées lorsque vous étiez connecté à votre cluster SSH, aux valeurs par défaut du cluster pour le groupe d'instances spécifié.

Considérations à prendre en compte lors de la reconfiguration d'un groupe d'instances

Actions de reconfiguration

Lorsque vous soumettez une demande de reconfiguration à l'aide de la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK, Amazon EMR vérifie le fichier de configuration existant sur le cluster. S'il existe des différences entre la configuration existante et le fichier que vous fournissez, Amazon EMR lance des actions de reconfiguration, redémarre certaines applications et réinitialise toutes les configurations modifiées manuellement, telles que les configurations que vous avez modifiées lorsque vous êtes connecté à votre cluster SSH, aux valeurs par défaut du cluster pour le groupe d'instances spécifié.

Note

Amazon EMR exécute certaines actions par défaut lors de chaque reconfiguration de groupe d'instances. Ces actions par défaut peuvent entrer en conflit avec les personnalisations de cluster que vous avez effectuées et entraîner des échecs de reconfiguration. Pour plus d'informations sur le dépannage des échecs de reconfiguration, consultez [Résoudre les problèmes de reconfiguration des groupes d'instances](#).

Amazon lance EMR également des actions de reconfiguration pour les classifications de configuration que vous spécifiez dans votre demande. Pour obtenir la liste complète de ces actions, consultez la section Classifications de configuration correspondant à la version d'Amazon EMR que vous utilisez. Par exemple, [Classifications de configuration 6.2.0](#).

Note

L'Amazon EMR Release Guide répertorie uniquement les actions de reconfiguration à partir des EMR versions 5.32.0 et 6.2.0 d'Amazon.

Interruption du service

Amazon EMR suit un processus continu pour reconfigurer les instances dans les groupes d'instances Task et Core. Seules 10 % des instances d'un groupe d'instances peuvent être modifiées et redémarrées à un moment donné. Ce processus est plus long à s'exécuter, mais il

permet de réduire le risque de défaillance potentielle des applications dans un cluster en cours d'exécution.

Pour exécuter YARN des tâches lors d'un YARN redémarrage, vous pouvez soit créer un EMR cluster Amazon avec plusieurs nœuds principaux, soit le `yarn.resourcemanager.recovery.enabled` définir `true` dans votre classification `yarn-site` de configuration. Pour plus d'informations sur l'utilisation de plusieurs nœuds maîtres, consultez la section [Haute disponibilité YARN ResourceManager](#).

Validation des applications

Amazon EMR vérifie que chaque application du cluster est en cours d'exécution après le redémarrage de la reconfiguration. Si une application n'est pas disponible, l'opération de reconfiguration globale échoue. Si une opération de reconfiguration échoue, Amazon EMR rétablit les paramètres de configuration par rapport à la version de travail précédente.


Note

Pour éviter tout échec de reconfiguration, nous vous recommandons de n'installer sur votre cluster que les applications que vous comptez utiliser. Nous vous recommandons également de vous assurer que toutes les applications de cluster fonctionnent correctement avant de soumettre une demande de reconfiguration.

Types de reconfiguration

Vous pouvez reconfigurer un groupe d'instances de deux manières suivantes :

- **Overwrite.** Méthode de reconfiguration par défaut et la seule disponible dans les EMR versions Amazon antérieures aux versions 5.35.0 et 6.6.0. Cette méthode de reconfiguration remplace sans distinction tous les fichiers du cluster par le nouvel ensemble de configuration soumis. La méthode efface toutes les modifications apportées aux fichiers de configuration en dehors de la API reconfiguration.
- **Fusionner.** Méthode de reconfiguration prise en charge pour les EMR versions 5.35.0 et 6.6.0 d'Amazon et versions ultérieures, sauf depuis la EMR console Amazon, où aucune version ne la prend en charge. Cette méthode de reconfiguration fusionne les configurations récemment soumises avec les configurations qui existent déjà sur le cluster. Cette option ajoute ou modifie uniquement les nouvelles configurations que vous soumettez. Il préserve les configurations existantes.

 Note

Amazon EMR continue de remplacer certaines configurations Hadoop essentielles dont elle a besoin pour garantir le bon fonctionnement du service.

Limites

Lorsque vous reconfigurez un groupe d'instances dans un cluster en cours d'exécution, tenez compte des limitations suivantes :

- **YARN** Les applications non opérationnelles peuvent échouer au redémarrage ou provoquer des problèmes de cluster, en particulier si les applications ne sont pas correctement configurées. Les clusters approchant le maximum de mémoire et CPU d'utilisation peuvent rencontrer des problèmes après le redémarrage. Cela est particulièrement vrai pour le groupe d'instances principal.
- Vous ne pouvez pas envoyer de demande de reconfiguration lorsqu'un groupe d'instances est redimensionné. Si une reconfiguration est initiée pendant le redimensionnement d'un groupe d'instances, la reconfiguration ne peut pas démarrer tant que le redimensionnement du groupe d'instances n'est pas terminé, et vice versa.
- Après avoir reconfiguré un groupe d'instances, Amazon EMR redémarre les applications pour permettre aux nouvelles configurations de prendre effet. Des échecs de tâches ou d'autres comportements inattendus des applications peuvent se produire si les applications sont utilisées durant la reconfiguration.
- Si la reconfiguration d'un groupe d'instances échoue, Amazon EMR rétablit les paramètres de configuration par rapport à la version de travail précédente. Si le processus de réinitialisation échoue, vous devez soumettre une nouvelle demande `ModifyInstanceGroup` pour récupérer le groupe d'instances à partir de l'état `SUSPENDED`.
- Les demandes de reconfiguration pour les classifications de configuration Phoenix ne sont prises en charge que dans les EMR versions 5.23.0 et ultérieures d'Amazon, et ne sont pas prises en charge dans les EMR versions 5.21.0 ou 5.22.0 d'Amazon.
- Les demandes de reconfiguration pour les classifications HBase de configuration ne sont prises en charge que dans les EMR versions 5.30.0 et ultérieures d'Amazon, et ne sont pas prises en charge dans les EMR versions 5.23.0 à 5.29.0 d'Amazon.

- Amazon EMR prend en charge les demandes de reconfiguration d'applications sur un EMR cluster Amazon comportant plusieurs nœuds principaux uniquement dans les EMR versions 5.27.0 et ultérieures d'Amazon.
- La reconfiguration de la `hdfs-encryption-zones` classification ou de l'une des classifications de KMS configuration Hadoop n'est pas prise en charge sur un EMR cluster Amazon comportant plusieurs nœuds principaux.
- Amazon ne prend EMR actuellement pas en charge certaines demandes de reconfiguration du planificateur de capacité qui nécessitent le redémarrage du YARN ResourceManager Par exemple, vous ne pouvez pas supprimer complètement une file d'attente.

Reconfigurer un groupe d'instances dans la console

Note

La EMR console Amazon ne prend pas en charge les reconfigurations de type Merge.

1. Ouvrez la EMR console Amazon à l'adresse <https://console.aws.amazon.com/emr>
2. Dans la liste des clusters sous Nom, sélectionnez le cluster actif que vous souhaitez reconfigurer.
3. Ouvrez la page des détails du cluster et accédez à l'onglet Configurations.
4. Dans la liste déroulante Filter (Filtre), sélectionnez le groupe d'instances que vous souhaitez reconfigurer.
5. Dans le menu déroulant Reconfigurer, choisissez Modifier dans le tableau ou Modifier dans JSON le fichier.
 - Modifier dans la table – Dans la table de classification de configuration, modifiez la propriété et la valeur des configurations existantes ou choisissez Ajouter une configuration pour fournir des classifications de configuration supplémentaires.
 - Modifier dans JSON le fichier - Entrez la configuration directement dans JSON le fichier ou utilisez une syntaxe abrégée (illustrée dans le texte en ombre). Sinon, fournissez un Amazon S3 URI pour un fichier contenant un JSON Configurations objet.

Note

La colonne Source de la table des classifications de configuration indique si la configuration est fournie lorsque vous créez un cluster ou lorsque vous spécifiez des configurations supplémentaires pour ce groupe d'instances. Vous pouvez modifier les configurations d'un groupe d'instances à partir de ces deux sources. Vous ne pouvez pas supprimer les configurations de cluster initiales, mais vous pouvez les remplacer pour un groupe d'instances.

Vous pouvez également ajouter ou modifier les classifications de configuration imbriquées directement dans la table. Par exemple, pour fournir une sous-classification `export` supplémentaire de `hadoop-env`, ajoutez une classification de configuration `hadoop.export` dans la table. Ensuite, indiquez une propriété et une valeur spécifiques pour cette classification.

6. (Facultatif) Sélectionnez Apply this configuration to all active instance groups (Appliquer cette configuration à tous les groupes d'instances actifs).
7. Enregistrez les Modifications.

Reconfigurez un groupe d'instances à l'aide du CLI

Utilisez la commande `modify-instance-groups` pour spécifier une nouvelle configuration pour un groupe d'instances dans un cluster en cours d'exécution.

Note

Dans les exemples suivants, remplacez `<j-2AL4XXXXXX5T9>` avec votre identifiant de cluster, et remplacez `<ig-1xxxxxxx9>` avec l'ID de votre groupe d'instances.

Exemple – Remplacer une configuration pour un groupe d'instances

L'exemple suivant fait référence à un JSON fichier de configuration appelé `instanceGroups.json` pour modifier la propriété du vérificateur de santé du YARN NodeManager disque pour un groupe d'instances.

1. Préparez votre classification de configuration et enregistrez-la comme `instanceGroups.json` dans le répertoire où vous allez exécuter la commande.

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.disk-health-checker.enable": "true",
          "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage": "100.0"
        },
        "Configurations": []
      }
    ]
  }
]
```

2. Exécutez la commande suivante.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Exemple – Ajoute une configuration à un groupe d'instances

Si vous voulez ajouter une configuration à un groupe d'instances, vous devez inclure toutes les configurations précédemment spécifiées pour ce groupe d'instances dans votre nouvelle demande `ModifyInstanceGroup`. Dans le cas contraire, les configurations précédemment spécifiées sont supprimées.

L'exemple suivant ajoute une propriété pour le vérificateur de mémoire YARN NodeManager virtuelle. La configuration inclut également les valeurs précédemment spécifiées pour le vérificateur de santé du YARN NodeManager disque afin que les valeurs ne soient pas remplacées.

1. Préparez le contenu suivant dans `instanceGroups.json` et enregistrez-le dans le même répertoire que celui où vous exécuterez la commande.

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.disk-health-checker.enable": "true",
          "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage": "100.0",
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        },
        "Configurations": []
      }
    ]
  }
]
```

2. Exécutez la commande suivante.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Exemple – Ajoutez une configuration à un groupe d'instances avec une reconfiguration de type Merge

Lorsque vous souhaitez utiliser la méthode de reconfiguration Overwrite par défaut pour ajouter une configuration, vous devez inclure toutes les configurations précédemment spécifiées pour ce groupe d'instances dans votre nouvelle demande ModifyInstanceGroup. Dans le cas contraire, Overwrite supprime les configurations que vous avez spécifiées précédemment. Vous n'avez pas besoin de le faire avec la reconfiguration de Merge. Vous devez plutôt vous assurer que votre demande inclut uniquement les nouvelles configurations.

L'exemple suivant ajoute une propriété pour le vérificateur de mémoire YARN NodeManager virtuelle. Comme il s'agit d'une reconfiguration de type Merge, elle ne remplace pas les valeurs précédemment spécifiées pour le vérificateur de santé du YARN NodeManager disque.

1. Préparez le contenu suivant dans `instanceGroups.json` et enregistrez-le dans le même répertoire que celui où vous exécuterez la commande.

```
[
  {
    "InstanceId": "<ig-1xxxxxxx9>",
    "ReconfigurationType": "MERGE",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        },
        "Configurations": []
      }
    ]
  }
]
```

2. Exécutez la commande suivante.

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Exemple – Supprimer une configuration pour un groupe d'instances

Pour supprimer une configuration pour un groupe d'instances, soumettez une nouvelle demande de reconfiguration excluant la configuration précédente.

Note

Vous pouvez uniquement remplacer la configuration de cluster initiale. Vous ne pouvez pas la supprimer.

Par exemple, pour supprimer la configuration du vérificateur de santé du YARN NodeManager disque de l'exemple précédent, soumettez-en une nouvelle `instanceGroups.json` avec le contenu suivant.

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        },
        "Configurations": []
      }
    ]
  }
]
```

Note

Pour supprimer toutes les configurations de votre dernière demande de reconfiguration, soumettez une demande de reconfiguration avec un ensemble de configurations vide. Par exemple,

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "Configurations": []
  }
]
```

Exemple – Reconfigurez et redimensionnez un groupe d'instances en une seule demande

L'exemple suivant JSON montre comment reconfigurer et redimensionner un groupe d'instances dans la même demande.

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
```

```

    "InstanceCount":5,
    "EC2InstanceIdsToTerminate":["i-123"],
    "ForceShutdown":true,
    "ShrinkPolicy":{
      "DecommissionTimeout":10,
      "InstanceResizePolicy":{
        "InstancesToTerminate":["i-123"],
        "InstancesToProtect":["i-345"],
        "InstanceTerminationTimeout":20
      }
    },
    "Configurations":[
      {
        "Classification":"yarn-site",
        "Configurations":[],
        "Properties":{
          "yarn.nodemanager.disk-health-checker.enable":"true",
          "yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-
percentage":"100.0"
        }
      }
    ]
  }
]

```

Reconfigurer un groupe d'instances à l'aide de Java SDK

Note

Dans les exemples suivants, remplacez `<j-2AL4XXXXXX5T9>` avec votre identifiant de cluster, et remplacez `<ig-1xxxxxxx9>` avec l'ID de votre groupe d'instances.

L'extrait de code suivant fournit une nouvelle configuration pour un groupe d'instances utilisant AWS SDK for Java.

```

AWSCredentials credentials = new BasicAWSCredentials("access-key", "secret-key");
AmazonElasticMapReduce emr = new AmazonElasticMapReduceClient(credentials);

Map<String,String> hiveProperties = new HashMap<String,String>();

```



```
hiveProperties.put("hive.join.emit.interval","1000");
hiveProperties.put("hive.merge.mapfiles","true");

Configuration configuration = new Configuration()
    .withClassification("hive-site")
    .withProperties(hiveProperties);

InstanceGroupModifyConfig igConfig = new InstanceGroupModifyConfig()
    .withInstanceId("<ig-1xxxxxxx9>")
    .withReconfigurationType("MERGE");
    .withConfigurations(configuration);

ModifyInstanceGroupsRequest migRequest = new ModifyInstanceGroupsRequest()
    .withClusterId("<j-2AL4XXXXXX5T9>")
    .withInstanceGroups(igConfig);

emr.modifyInstanceGroups(migRequest);
```

L'extrait de code suivant supprime une configuration précédemment spécifiée pour un groupe d'instances en fournissant un tableau vide de configurations.

```
List<Configuration> configurations = new ArrayList<Configuration>();

InstanceGroupModifyConfig igConfig = new InstanceGroupModifyConfig()
    .withInstanceId("<ig-1xxxxxxx9>")
    .withConfigurations(configurations);

ModifyInstanceGroupsRequest migRequest = new ModifyInstanceGroupsRequest()
    .withClusterId("<j-2AL4XXXXXX5T9>")
    .withInstanceGroups(igConfig);

emr.modifyInstanceGroups(migRequest);
```

Résoudre les problèmes de reconfiguration des groupes d'instances

Si le processus de reconfiguration d'un groupe d'instances échoue, Amazon annule la EMR reconfiguration et enregistre un message d'échec à l'aide d'un événement Amazon. CloudWatch L'événement fournit un bref résumé de l'échec de la reconfiguration. Il répertorie les instances pour lesquelles la reconfiguration a échoué et les messages d'échec correspondants. Voici un exemple de message d'échec.

```
The reconfiguration operation for instance group ig-1xxxxxxx9 in Amazon EMR
cluster j-2AL4XXXXXX5T9 (ExampleClusterName)
failed at 2021-01-01 00:00 UTC and took 2 minutes to fail. Failed configuration version
is example12345.
Failure message: Instance i-xxxxxxx1, i-xxxxxxx2, i-xxxxxxx3 failed with message "This
is an example failure message".
```

Pour recueillir davantage de données sur un échec de reconfiguration, vous pouvez consulter les journaux de provisionnement des nœuds. Cela est particulièrement utile lorsque vous recevez un message comme celui-ci.

```
i-xxxxxxx1 failed with message "Unable to complete transaction and some changes were
applied."
```

On the node

Pour accéder aux journaux de provisionnement des nœuds en vous connectant à un nœud

1. SSH à utiliser pour se connecter au nœud sur lequel la reconfiguration a échoué. Pour obtenir des instructions, consultez [Connect to your Linux instance](#) dans le guide de EC2 l'utilisateur Amazon pour les instances Linux.
2. Accédez au répertoire suivant, qui contient les fichiers journaux de provisionnement des nœuds.

```
/mnt/var/log/provision-node/
```

3. Ouvrez le sous-répertoire `reports` et recherchez le rapport de provisionnement des nœuds pour votre reconfiguration. Le `reports` répertoire organise les journaux par numéro de version de reconfiguration, identifiant unique universel (UUID), adresse IP de l'EC2instance Amazon et horodatage. Chaque rapport est un YAML fichier compressé qui contient des informations détaillées sur le processus de reconfiguration.

Vous trouverez ci-dessous un exemple de nom de fichier et de chemin d'accès.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxx/ip-10-73-xxx-
xxx.ec2.internal/202104061715.yaml.gz
```

4. Vous pouvez examiner un rapport à l'aide d'un visualiseur de fichiers tel que `zless`, comme dans l'exemple suivant.

```
zless 202104061715.yaml.gz
```

Amazon S3

Pour accéder aux journaux de provisionnement des nœuds à l'aide d'Amazon S3

1. Connectez-vous à la console Amazon S3 AWS Management Console et ouvrez-la à l'adresse <https://console.aws.amazon.com/s3/>.
2. Ouvrez le compartiment Amazon S3 que vous avez spécifié lorsque vous avez configuré le cluster pour archiver les fichiers journaux.
3. Accédez au dossier suivant, qui contient les fichiers journaux de provisionnement des nœuds :

```
DOC-EXAMPLE-BUCKET/elasticmapreduce/<cluster id>/node/<instance id>/provision-  
node/
```

4. Ouvrez le dossier `reports` et recherchez le rapport de provisionnement des nœuds pour votre reconfiguration. Le `reports` dossier organise les journaux par numéro de version de reconfiguration, identifiant unique universel (UUID), adresse IP de l'EC2instance Amazon et horodatage. Chaque rapport est un YAML fichier compressé qui contient des informations détaillées sur le processus de reconfiguration.

Vous trouverez ci-dessous un exemple de nom de fichier et de chemin d'accès.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxxxx/ip-10-73-xxx-  
xxx.ec2.internal/202104061715.yaml.gz
```

5. Pour consulter un fichier journal, vous pouvez le télécharger depuis Amazon S3 sur votre machine locale sous forme de fichier texte. Pour obtenir des instructions, consultez la section [Téléchargement d'un objet](#).

Chaque fichier journal contient un rapport de provisionnement détaillé pour la reconfiguration associée. Pour trouver les informations relatives aux messages d'erreur, vous pouvez rechercher le niveau de journalisation `err` d'un rapport. Le format du rapport dépend de la version d'Amazon EMR de votre cluster.

L'exemple suivant montre les informations d'erreur relatives aux EMR versions d'Amazon antérieures aux versions 5.32.0 et 6.2.0.

```
- !ruby/object:Puppet::Util::Log
  level: !ruby/sym err
  tags:
    - err
  message: "Example detailed error message."
  source: Puppet
  time: 2021-01-01 00:00:00.000000 +00:00
```

Les EMR versions 5.32.0 et 6.2.0 et ultérieures d'Amazon utilisent plutôt le format suivant.

```
- level: err
  message: 'Example detailed error message.'
  source: Puppet
  tags:
    - err
  time: '2021-01-01 00:00:00.000000 +00:00'
  file:
  line:
```

Stocker les données de configuration sensibles dans AWS Secrets Manager

Les API opérations de EMR description et de liste Amazon qui émettent des données de configuration personnalisées (telles que `DescribeCluster` et `ListInstanceGroups`) le font en texte clair. Amazon EMR s' AWS Secrets Manager y intègre afin que vous puissiez stocker vos données dans Secrets Manager et utiliser le secret ARN dans vos configurations. Ainsi, vous ne transmettez pas de données de configuration sensibles à Amazon EMR en texte brut et ne les exposez pas à des données externes APIs. Si vous indiquez qu'une paire clé-valeur contient un ARN pour un secret stocké dans Secrets Manager, Amazon EMR récupère ce secret lorsqu'il envoie les données de configuration au cluster. Amazon EMR n'envoie pas l'annotation lorsqu'il utilise un outil externe APIs pour afficher la configuration.

Créer un secret

Pour créer un secret, suivez les étapes de la section [Création d'un secret AWS Secrets Manager](#) dans le Guide de l'utilisateur AWS Secrets Manager . Dans l'étape 3, vous devez choisir le champ de texte en clair pour saisir votre valeur sensible.

Notez que si Secrets Manager autorise un secret à contenir jusqu'à 65 536 octets, Amazon EMR limite la longueur combinée de la clé de propriété (à l'exception de l'annotation) et de la valeur secrète récupérée à 1 024 caractères.

Accordez EMR à Amazon l'accès pour récupérer le secret

Amazon EMR utilise un rôle IAM de service pour provisionner et gérer des clusters pour vous. Le rôle de service d'Amazon EMR définit les actions autorisées pour Amazon EMR lorsqu'il fournit des ressources et exécute des tâches de niveau de service qui ne sont pas effectuées dans le contexte d'une EC2 instance Amazon exécutée au sein d'un cluster. Pour plus d'informations sur les rôles de service, consultez [Rôle de service pour Amazon EMR \(EMRrôle\)](#) et [Personnalisation IAM des rôles](#).

Pour permettre EMR à Amazon de récupérer la valeur secrète auprès de Secrets Manager, ajoutez la déclaration de politique suivante à votre EMR rôle Amazon lorsque vous lancez votre cluster.

```
{
  "Sid":"AllowSecretsRetrieval",
  "Effect":"Allow",
  "Action":"secretsmanager:GetSecretValue",
  "Resource":[
    "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  ]
}
```

Si vous créez le secret avec une clé gérée par le client AWS KMS key, vous devez également ajouter une `kms:Decrypt` autorisation au EMR rôle Amazon pour la clé que vous utilisez. Pour plus d'informations, consultez [Authentification et contrôle d'accès pour AWS Secrets Manager](#) dans le Guide de l'utilisateur AWS Secrets Manager .

Utiliser le secret dans une classification de configuration

Vous pouvez ajouter l'EMR .secret@annotation à n'importe quelle propriété de configuration pour indiquer que sa paire clé-valeur contient un ARN pour un secret stocké dans Secrets Manager.

L'exemple suivant montre comment fournir un secret ARN dans une classification de configuration :

```
{
  "Classification":"core-site",
  "Properties":{
    "presto.s3.access-key":"<sensitive-access-key>",
    "EMR.secret@presto.s3.secret-key":"arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  }
}
```

Lorsque vous créez votre cluster et soumettez votre configuration annotée, Amazon EMR valide les propriétés de configuration. Si votre configuration est valide, Amazon EMR supprime l'annotation de la configuration et récupère le secret auprès de Secrets Manager pour créer la configuration réelle avant de l'appliquer au cluster :

```
{
  "Classification":"core-site",
  "Properties":{
    "presto.s3.access-key":"<sensitive-access-key>",
    "presto.s3.secret-key":"<my-secret-key-retrieved-from-Secrets-Manager>"
  }
}
```

Lorsque vous appelez une action telle que `DescribeCluster`, Amazon EMR renvoie la configuration actuelle de l'application sur le cluster. Si une propriété de configuration d'application est marquée comme contenant un secret ARN, la configuration d'application renvoyée par l'`DescribeCluster` appel contient la valeur secrète ARN et non la valeur secrète. Cela garantit que la valeur secrète n'est visible que sur le cluster :

```
{
  "Classification":"core-site",
  "Properties":{
    "presto.s3.access-key":"<sensitive-access-key>",
    "presto.s3.secret-key":"arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  }
}
```

Mettre à jour la valeur d'un secret

Amazon EMR récupère la valeur secrète d'une configuration annotée chaque fois que le groupe d'instances attaché démarre, se reconfigure ou se redimensionne. Vous pouvez utiliser Secrets Manager pour modifier la valeur d'un secret utilisé dans la configuration d'un cluster en cours d'exécution. Dans ce cas, vous pouvez envoyer une demande de reconfiguration à chaque groupe d'instances pour lequel vous souhaitez recevoir la valeur mise à jour. Pour plus d'informations sur la façon de reconfigurer un groupe d'instances et sur les éléments à prendre en compte lorsque vous le faites, consultez [Reconfigurer un groupe d'instances dans un cluster en cours d'exécution](#).

Configuration des applications pour l'utilisation d'une machine virtuelle Java spécifique

Les EMR versions d'Amazon proposent différentes versions par défaut de Java Virtual Machine (JVM). Cette page explique la JVM prise en charge des différentes versions et applications.

Considérations

Pour plus d'informations sur les versions Java prises en charge pour les applications, consultez les pages relatives aux applications dans le [Amazon EMR Release Guide](#).

- Amazon EMR ne prend en charge que l'exécution d'une seule version d'exécution dans un cluster, et ne prend pas en charge l'exécution de différents nœuds ou applications sur différentes versions d'exécution sur le même cluster.
- Pour Amazon EMR 7.x, la machine virtuelle Java par défaut (JVM) est Java 17 pour les applications compatibles avec Java 17, à l'exception d'Apache Livy. Pour plus d'informations sur les JDK versions prises en charge pour les applications, consultez la page de publication correspondante dans le Amazon EMR Release Guide.
- À partir d'Amazon EMR 7.1.0, Flink prend en charge Java 17 et est configuré sur Java 17 par défaut. Pour utiliser une version différente de l'environnement d'exécution Java, remplacez les paramètres définis dans `flink-conf`. Pour plus d'informations sur la configuration de Flink pour utiliser Java 8 ou Java 11, voir [Configurer Flink pour fonctionner avec Java 11](#).
- Pour Amazon EMR 5.x et 6.x, la machine virtuelle Java par défaut (JVM) est Java 8.
 - Pour les EMR versions 6.12.0 et supérieures d'Amazon, certaines applications prennent également en charge Java 11 et 17.

- Pour les EMR versions 6.9.0 et supérieures d'Amazon, Trino prend en charge Java 17 par défaut. Pour plus d'informations sur Java 17 avec Trino, consultez [Trino updates to Java 17](#) sur le blog de Trino.

Tenez compte des considérations spécifiques à l'application suivantes lorsque vous choisissez votre version d'exécution :

Remarques de configuration Java spécifiques à l'application

Application	Notes sur la configuration de Java
Spark	<p>Pour exécuter Spark avec une version de Java autre que celle par défaut, vous devez configurer à la fois Spark et Hadoop. Pour obtenir des exemples, consultez Remplacez le JVM.</p> <ul style="list-style-type: none"> • Configurez <code>JAVA_HOME</code> dans <code>spark-env</code> pour mettre à jour le runtime Java des processus de l'instance principale. Par exemple, <code>spark-submit</code>, <code>spark-shell</code> et Spark History Server. • Modifiez la configuration Hadoop pour mettre à jour le runtime Java des exécuteurs Spark et du YARN ApplicationMaster
Étincelle RAPIDS	Vous pouvez exécuter RAPIDS avec la version Java configurée pour Spark.
Iceberg	Vous pouvez exécuter Iceberg avec la version Java configurée de l'application qui l'utilise.
Delta	Vous pouvez exécuter Delta avec la version Java configurée de l'application qui l'utilise.
Hudi	Vous pouvez exécuter Hudi avec la version Java configurée de l'application qui l'utilise.

Application	Notes sur la configuration de Java
Hadoop	Pour mettre à jour le JVM pour Hadoop, modifiez <code>hadoop-env</code> . Pour obtenir des exemples, consultez Remplacez le JVM .
Hive	Pour définir la version Java sur 11 ou 17 pour Hive, configurez le JVM paramètre Hadoop sur la version Java que vous souhaitez utiliser.
HBase	Pour mettre à jour le JVM formulaire HBase, modifiez <code>lebase-env</code> . Par défaut, Amazon EMR définit le en HBase JVM fonction de la JVM configuration de Hadoop, sauf si vous remplacez les paramètres dans <code>hbase-env</code> . Pour obtenir des exemples, consultez Remplacez le JVM .
Flink	Pour mettre à jour le JVM for Flink, <code>flink-conf</code> modifiez-le. Par défaut, Amazon EMR définit le Flink en JVM fonction de la JVM configuration de Hadoop, sauf si vous remplacez les paramètres dans <code>flink-conf</code> . Pour plus d'informations, consultez Configurer Flink pour qu'il fonctionne avec Java 11 .
Oozie	Pour configurer Oozie pour qu'il fonctionne sur Java 11 ou 17, configurez Oozie Server, le Oozie Launcher AM Launcher AM et modifiez les configurations de vos exécutables et tâches côté client. Vous pouvez également configurer <code>EmbeddedOozieServer</code> pour l'exécuter sur Java 17. Pour plus d'informations, consultez Configurer la version Java pour Oozie .
Pig	Pig ne supporte que Java 8. Vous ne pouvez pas utiliser Java 11 ou 17 avec Hadoop et exécuter Pig sur le même cluster.

Remplacez le JVM

Pour remplacer le JVM paramètre d'une EMR version d'Amazon (par exemple, pour utiliser Java 17 avec un cluster utilisant la EMR version 6.12.0 d'Amazon), fournissez le JAVA_HOME paramètre correspondant à sa classification d'environnement, qui s'applique à toutes *application*-env les applications sauf Flink. Pour Flink, la classification de l'environnement est `flink-conf`. Pour connaître les étapes de configuration de l'environnement d'exécution Java avec Flink, consultez [Configurer Flink pour qu'il fonctionne avec Java 11](#).

Rubriques

- [Remplacer le JVM paramètre par Apache Spark](#)
- [Remplacer le JVM paramètre par Apache HBase](#)
- [Remplacez le JVM paramètre par Apache Hadoop et Hive](#)

Remplacer le JVM paramètre par Apache Spark

Lorsque vous utilisez Spark avec les EMR versions 6.12 et supérieures d'Amazon, si vous écrivez un pilote à soumettre en mode cluster, le pilote utilise Java 8, mais vous pouvez configurer l'environnement de telle sorte que les exécuteurs utilisent Java 11 ou 17. Et lorsque vous utilisez Spark avec des EMR versions d'Amazon inférieures à la version 5.x et que vous écrivez un pilote à soumettre en mode cluster, le pilote utilise Java 7. Vous pouvez toutefois configurer l'environnement de manière à ce que les exécuteurs utilisent Java 8.

Pour remplacer le JVM for Spark, nous vous recommandons de définir à la fois les classifications Hadoop et Spark.

```
{
  "Classification": "hadoop-env",
  "Configurations": [
    {
      "Classification": "export",
      "Configurations": [],
      "Properties": {
        "JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
      }
    }
  ],
  "Properties": {}
},
```

```

{
  "Classification": "spark-env",
    "Configurations": [
      {
        "Classification": "export",
          "Configurations": [],
          "Properties": {
            "JAVA_HOME": "/usr/lib/jvm/java-1.8.0"
          }
        }
      ],
      "Properties": {}
    }
  }

```

Remplacer le JVM paramètre par Apache HBase

Pour configurer HBase pour utiliser Java 11, vous pouvez définir la configuration suivante lorsque vous lancez le cluster.

```

[
  {
    "Classification": "hbase-env",
    "Configurations": [
      {
        "Classification": "export",
          "Configurations": [],
          "Properties": {
            "JAVA_HOME": "/usr/lib/jvm/jre-11"
          }
        }
      ],
      "Properties": {}
    }
  ]

```

Remplacez le JVM paramètre par Apache Hadoop et Hive

L'exemple suivant montre comment définir la JVM version 17 pour Hadoop et Hive.

```

[
  {

```

```

    "Classification": "hadoop-env",
    "Configurations": [
      {
        "Classification": "export",
        "Configurations": [],
        "Properties": {
          "JAVA_HOME": "/usr/lib/jvm/jre-17"
        }
      }
    ],
    "Properties": {}
  }
]

```


Ports de service

Les ports suivants sont YARN des ports HDFS de service. Ces paramètres reflètent les valeurs par défaut de Hadoop. D'autres services d'applications sont hébergés sur les ports par défaut, sauf indication contraire. Pour de plus amples informations, veuillez consulter la documentation du projet de l'application.

Paramètres de port pour YARN et HDFS

Paramètre	Nom d'hôte/Port
fs.default.name	par défaut (hdfs:// <i>emrDeterminedIP</i> :8020)
dfs.datanode.address	par défaut (0.0.0.0:50010)
dfs.datanode.http.address	par défaut (0.0.0.0:50075)
dfs.datanode.https.address	par défaut (0.0.0.0:50475)
dfs.datanode.ipc.address	par défaut (0.0.0.0:50020)
dfs.http.address	par défaut (0.0.0.0:50070)
dfs.https.address	par défaut (0.0.0.0:50470)
dfs.secondary.http.address	par défaut (0.0.0.0:50090)

Paramètre	Nom d'hôte/Port
<code>yarn.nodemanager.address</code>	par défaut (<code>\${yarn.nodemanager.hostname}:0</code>)
<code>yarn.nodemanager.localizer.address</code>	par défaut (<code>\${yarn.nodemanager.hostname}:8040</code>)
<code>yarn.nodemanager.webapp.address</code>	par défaut (<code>\${yarn.nodemanager.hostname}:8042</code>)
<code>yarn.resourcemanager.address</code>	par défaut (<code>\${yarn.resourcemanager.hostname}:8032</code>)
<code>yarn.resourcemanager.admin.address</code>	par défaut (<code>\${yarn.resourcemanager.hostname}:8033</code>)
<code>yarn.resourcemanager.resource-tracker.address</code>	par défaut (<code>\${yarn.resourcemanager.hostname}:8031</code>)
<code>yarn.resourcemanager.scheduler.address</code>	par défaut (<code>\${yarn.resourcemanager.hostname}:8030</code>)
<code>yarn.resourcemanager.webapp.address</code>	par défaut (<code>\${yarn.resourcemanager.hostname}:8088</code>)
<code>yarn.web-proxy.address</code>	par défaut (no-value)
<code>yarn.resourcemanager.hostname</code>	<i>emrDeterminedIP</i>

 Note

Le terme *emrDeterminedIP* est une adresse IP générée par le plan de EMR contrôle Amazon. Dans la version la plus récente, cette convention a été supprimée, sauf pour les paramètres `yarn.resourcemanager.hostname` et `fs.default.name`.

Utilisateurs de l'application

Les applications exécutent des processus comme s'il s'agissait de leur propre utilisateur. Par exemple, Hive JVMs s'exécute en tant qu'utilisateurhive, MapReduce JVMs exécute en tant qu'utilisateurmapred, etc. Ceci est illustré dans l'exemple d'état des processus suivant.

```

USER      PID %CPU %MEM    VSZ   RSS TTY      STAT START   TIME COMMAND
hive      6452  0.2  0.7 853684 218520 ?        S1   16:32   0:13 /usr/lib/jvm/
java-openjdk/bin/java -Xmx256m -Dhive.log.dir=/var/log/hive -Dhive.log.file=hive-
metastore.log -Dhive.log.threshold=INFO -Dhadoop.log.dir=/usr/lib/hadoop
hive      6557  0.2  0.6 849508 202396 ?        S1   16:32   0:09 /usr/lib/jvm/java-
openjdk/bin/java -Xmx256m -Dhive.log.dir=/var/log/hive -Dhive.log.file=hive-server2.log
-Dhive.log.threshold=INFO -Dhadoop.log.dir=/usr/lib/hadoop/1
hbase     6716  0.1  1.0 1755516 336600 ?        S1   Jun21   2:20 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_master -XX:OnOutOfMemoryError=kill -9 %p -Xmx1024m -ea -XX:
+UseConcMarkSweepGC -XX:+CMSIncrementalMode -Dhbase.log.dir=/var/
hbase     6871  0.0  0.7 1672196 237648 ?        S1   Jun21   0:46 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_thrift -XX:OnOutOfMemoryError=kill -9 %p -Xmx1024m -ea -XX:
+UseConcMarkSweepGC -XX:+CMSIncrementalMode -Dhbase.log.dir=/var/
hdfs      7491  0.4  1.0 1719476 309820 ?        S1   16:32   0:22 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_namenode -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-hdfs -
Dhadoop.log.file=hadoop-hdfs-namenode-ip-10-71-203-213.log -Dhadoo
yarn      8524  0.1  0.6 1626164 211300 ?        S1   16:33   0:05 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_proxyserver -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-yarn -
Dyarn.log.dir=/var/log/hadoop-yarn -Dhadoop.log.file=yarn-yarn-
yarn      8646  1.0  1.2 1876916 385308 ?        S1   16:33   0:46 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_resourcemanager -Xmx1000m -Dhadoop.log.dir=/var/log/hadoop-yarn
-Dyarn.log.dir=/var/log/hadoop-yarn -Dhadoop.log.file=yarn-y
mapred    9265  0.2  0.8 1666628 260484 ?        S1   16:33   0:12 /usr/lib/jvm/java-
openjdk/bin/java -Dproc_historyserver -Xmx1000m -Dhadoop.log.dir=/usr/lib/hadoop/logs -
Dhadoop.log.file=hadoop.log -Dhadoop.home.dir=/usr/lib/hadoop

```

Vérification des dépendances à l'aide du référentiel d'artefacts d'Amazon EMR

Vous pouvez utiliser le référentiel d'artefacts Amazon EMR pour générer le code de travail Apache Hive et Apache Hadoop sur les versions exactes des bibliothèques et dépendances qui sont disponibles avec des versions Amazon EMR spécifiques, en commençant par la version 5.18.0 d'Amazon EMR . La création contre les artefacts Amazon EMR dans le référentiel contribue à éviter des problèmes de chemin d'accès de classe d'exécution en assurant que les versions des bibliothèques contre lesquelles la tâche est construite sont exactement les mêmes versions fournies au moment de l'exécution sur le cluster. Actuellement, les artefacts Amazon EMR ne sont disponibles que pour les builds Maven.

Pour accéder à l'artefact de référentiel, ajoutez l'URL du référentiel à votre fichier de paramètres Maven ou à un fichier de configuration pom.xml d'un projet spécifique. Vous pouvez alors spécifier les dépendances de votre configuration de projet. Pour les versions de dépendances, utilisez la version répertoriée sous Versions de composant pour la version souhaitée sur [Versions publiées par Amazon EMR 5.x](#). Par exemple, les versions des composants pour la dernière version Amazon EMR sont disponibles à l'adresse [the section called "Versions des composants 5.36.2"](#). Si un artefact pour votre projet n'est pas répertorié sous Versions de composant, spécifiez la version qui est répertoriée pour Hive et Hadoop dans cette version. Par exemple, pour les composants Hadoop dans la version 5.18.0 d'Amazon EMR, la version est 2.8.4-amzn-1.

L'URL de référentiel d'artefact dispose de la syntaxe suivante :

```
https://s3-endpoint/region-ID-emr-artifacts/emr-release-label/repos/maven/
```

- *s3-endpoint* est le point de terminaison Amazon Simple Storage Service (Amazon S3) de la région pour le référentiel et *region-ID* est la région correspondante. Par exemple : `s3.us-west-1.amazonaws.com` et `us-west-1`. Pour plus d'informations, consultez les points de terminaison Amazon S3 dans la Référence générale d'Amazon Web Services. Il n'y a pas de différence d'artefacts entre les régions. Cela vous permet de spécifier la région la plus pratique pour votre environnement de développement.
- *emr-release-label* est le label de publication du cluster Amazon EMR qui exécutera votre code. Les étiquettes de version respectent le format `emr-x.x.x`, par exemple, `emr-5.36.2`. Une série de versions EMR peut inclure plusieurs versions. Par exemple, si vous utilisez EMR version

5.24.1, utilisez la première étiquette de version EMR de la série 5.24, `emr-5.24.0`, dans l'URL du référentiel d'artefacts :

```
https://s3-endpoint/region-ID-emr-artifacts/emr-5.24.0/repos/maven/
```

Exemple Configuration pour Maven pom.xml

L'exemple de pom.xml ci-dessous configure un projet Maven pour se créer contre les artefacts Apache Hadoop et Apache Hive `emr-5.18.0`, à l'aide du référentiel d'artefacts dans `us-west-1`. Les versions d'instantanés ne sont pas disponibles dans le référentiel d'artefacts. Par conséquent, les instantanés sont désactivés dans le pom.xml. Des ellipses (...) dans l'exemple ci-dessous indiquent l'omission d'autres paramètres de configuration. Ne les copiez pas dans votre projet Maven.

```
<project>
...
<repositories>
...
<repository>
  <id>emr-5.18.0-artifacts</id>
  <name>EMR 5.18.0 Releases Repository</name>
  <releases>
    <enabled>true</enabled>
  </releases>
  <snapshots>
    <enabled>false</enabled>
  </snapshots>
  <url>https://s3.us-west-1.amazonaws.com/us-west-1-emr-artifacts/emr-5.18.0/repos/
maven/</url>
</repository>
...
</repositories>
...
<dependencies>
...
<dependency>
  <groupId>org.apache.hive</groupId>
  <artifactId>hive-exec</artifactId>
  <version>2.3.3-amzn-2</version>
</dependency>
<dependency>
  <groupId>org.apache.hadoop</groupId>
```



```
<artifactId>hadoop-common</artifactId>
<version>2.8.4-amzn-1</version>
</dependency>
...
</dependencies>

</project>
```

EMR Système de fichiers (EMRFS)

Le système de EMR fichiers (EMRFS) est une implémentation HDFS que tous les EMR clusters Amazon utilisent pour lire et écrire des fichiers ordinaires EMR directement depuis Amazon vers Amazon S3. EMRFS permet de stocker des données persistantes dans Amazon S3 pour une utilisation avec Hadoop, tout en fournissant des fonctionnalités telles que le chiffrement des données.

Le chiffrement des données vous permet de chiffrer des objets que EMRFS écrivent sur Amazon S3 et de EMRFS travailler avec des objets chiffrés dans Amazon S3. Si vous utilisez la EMR version 4.8.0 ou ultérieure d'Amazon, vous pouvez utiliser des configurations de sécurité pour configurer le chiffrement EMRFS des objets dans Amazon S3, ainsi que d'autres paramètres de chiffrement. Pour plus d'informations, consultez [Options de chiffrement](#). Si vous utilisez une version antérieure d'Amazon EMR, vous pouvez configurer manuellement les paramètres de chiffrement. Pour plus d'informations, consultez [Spécification du chiffrement Amazon S3 à l'aide EMRFS des propriétés](#).

Amazon S3 offre une grande cohérence en lecture après écriture pour tous GETPUT, ainsi que pour les LIST opérations dans l'ensemble. Régions AWS Cela signifie que ce que vous utilisez pour écrire EMRFS correspond à ce que vous allez lire sur Amazon S3, sans aucun impact sur les performances. Pour plus d'informations, consultez [Modèle de cohérence de données Amazon S3](#).

Lorsque vous utilisez la EMR version 5.10.0 ou ultérieure d'Amazon, vous pouvez utiliser différents IAM rôles pour les EMRFS demandes adressées à Amazon S3 en fonction des utilisateurs du cluster, des groupes ou de l'emplacement des EMRFS données dans Amazon S3. Pour plus d'informations, consultez [Configurer IAM les rôles pour les EMRFS demandes adressées à Amazon S3](#).

Warning

Avant d'activer l'exécution spéculative pour les EMR clusters Amazon exécutant des tâches Apache Spark, veuillez consulter les informations suivantes.

EMRFS inclut le committer EMRFS optimisé pour S3, une OutputCommitter implémentation optimisée pour écrire des fichiers sur Amazon S3 lors de l'utilisation. EMRFS [Si vous activez la fonctionnalité d'exécution spéculative d'Apache Spark avec des applications qui écrivent des données dans Amazon S3 et n'utilisent pas le validateur EMRFS optimisé pour S3, vous pouvez rencontrer les problèmes d'exactitude des données décrits dans la section -10063.](#) [SPARK](#) Cela peut se produire si vous utilisez des EMR versions d'Amazon antérieures à la EMR version 5.19 d'Amazon, ou si vous écrivez des fichiers dans Amazon S3 avec des formats tels que ORC et CSV. Ces formats ne sont pas pris en charge par le validateur

EMRFS optimisé pour S3. Pour une liste complète des exigences relatives à l'utilisation du validateur EMRFS optimisé S3, voir [Exigences relatives au validateur optimisé S3](#). EMRFS l'écriture directe est généralement utilisée lorsque le validateur EMRFS optimisé pour S3 n'est pas pris en charge, par exemple lors de l'écriture de ce qui suit :

- Un format de sortie autre que Parquet, tel que ORC le texte.
- Fichiers Hadoop utilisant le Spark. RDD API
- Parquet avec Hive. SerDe Consultez [Conversion de la table Hive Metastore Parquet](#).

EMRFS l'écriture directe n'est pas utilisée dans les scénarios suivants :

- Lorsque le validateur EMRFS optimisé pour S3 est activé. Voir les [exigences relatives au transmetteur EMRFS optimisé pour S3](#).
- Lors de l'écriture de partitions dynamiques `partitionOverwriteMode` définies sur dynamique.
- Lorsque vous écrivez sur des emplacements de partition personnalisés, tels que des emplacements non conformes à la convention d'emplacement de partition par défaut de Hive.
- Lorsque vous utilisez des systèmes de fichiers autres que EMRFS, par exemple, l'écriture HDFS ou l'utilisation du système de fichiers S3A.

Pour déterminer si votre application utilise l'écriture directe dans Amazon EMR 5.14.0 ou version ultérieure, activez la journalisation `SparkINFO`. Si une ligne de journal contenant le texte « Direct Write : ENABLED » est présente dans les journaux du pilote Spark ou dans les journaux des conteneurs de l'exécuteur Spark, cela signifie que votre application Spark a écrit en écriture directe.

Par défaut, l'exécution spéculative est activée OFF sur AmazonEMRclusters. Nous vous recommandons vivement de ne pas activer l'exécution spéculative si l'une ou l'autre de ces conditions est remplie :

- Vous écrivez des données sur Amazon S3.
- Les données sont écrites dans un format autre qu'Apache Parquet ou dans le format Apache Parquet sans utiliser le validateur EMRFS optimisé pour S3.

Si vous activez l'exécution spéculative de Spark et que vous écrivez des données sur Amazon S3 à l'aide de l'écriture EMRFS directe, vous risquez de subir des pertes de données

intermittentes. Lorsque vous écrivez des données ou que vous écrivez des données dans Parquet à l'aide du EMRFS validateur optimisé pour S3, Amazon n'utilise pas l'écriture directe et ce problème ne se produit pas. HDFS

Si vous devez écrire des données dans des formats utilisant l'écriture EMRFS directe de Spark vers Amazon S3 et utiliser l'exécution spéculative, nous vous recommandons d'écrire HDFS puis de transférer les fichiers de sortie vers Amazon S3 à l'aide de S3DistCP.

Rubriques

- [Vue cohérente](#)
- [Autoriser l'accès aux EMRFS données dans Amazon S3](#)
- [Gestion du point de AWS Security Token Service terminaison par défaut](#)
- [Spécification du chiffrement Amazon S3 à l'aide EMRFS des propriétés](#)

Vue cohérente

Warning

Le 1er juin 2023, le support standard pour les futures EMR versions d'Amazon expirera pour les vues EMRFS cohérentes. EMRFSune vue cohérente continuera de fonctionner pour les versions existantes.

Avec la sortie d'Amazon S3 Strong read-after-write Cohérence le 1er décembre 2020, vous n'avez plus besoin d'utiliser EMRFS une vue cohérente (EMRFSCV) avec vos EMR clusters Amazon. EMRFSLe CV est une fonctionnalité facultative qui permet aux EMR clusters Amazon de vérifier la liste et read-after-write la cohérence des objets Amazon S3. Lorsque vous créez un cluster et que le EMRFS CV est activé, Amazon EMR crée une base de données Amazon DynamoDB pour stocker les métadonnées des objets qu'il utilise pour suivre la liste read-after-write et la cohérence des objets S3. Vous pouvez désormais désactiver EMRFS CV et supprimer la base de données DynamoDB qu'il utilise afin de ne pas générer de coûts supplémentaires. Les procédures suivantes expliquent comment vérifier la fonctionnalité CV, la désactiver et supprimer la base de données DynamoDB qu'elle utilise.

Pour vérifier si vous utilisez la fonction EMRFS CV

1. Accédez à l'onglet Configuration. Si votre cluster possède la configuration suivante, il utilise EMRFS CV.

```
Classification=emrfs-site,Property=fs.s3.consistent,Value=true
```

2. Vous pouvez également utiliser le AWS CLI pour décrire votre cluster à l'aide du [describe-clusterAPI](#). Si la sortie contient `fs.s3.consistent: true`, votre cluster utilise EMRFS CV.

Pour désactiver le EMRFS CV sur vos EMR clusters Amazon

Pour désactiver la fonction EMRFS CV, utilisez l'une des trois options suivantes. Vous devez tester ces options dans votre environnement de test avant de les appliquer à vos environnements de production.

1. Pour arrêter votre cluster existant et en démarrer un nouveau sans options EMRFS CV.
 - a. Avant d'arrêter votre cluster, assurez-vous de sauvegarder vos données et d'en informer les utilisateurs.
 - b. Pour arrêter votre cluster, suivez les instructions de la section [Résilier un cluster](#).
 - c. Si vous utilisez la EMR console Amazon pour créer un nouveau cluster, accédez aux options avancées. Dans la section Modifier les paramètres du logiciel, désélectionnez l'option permettant d'activer le EMRFS CV. Si la case à cocher pour EMRFS une vue cohérente est disponible, désactivez-la.
 - d. Si vous avez l'habitude de AWS CLI créer un nouveau cluster avec le [create-clusterAPI](#), n'utilisez pas l'option `--emrfs`, qui active le EMRFS CV.
 - e. Si vous utilisez un SDK ou AWS CloudFormation pour créer un nouveau cluster, n'utilisez aucune des configurations répertoriées dans [Configurer une vue cohérente](#).
2. Pour cloner un cluster et supprimer un EMRFS CV
 - a. Dans la EMR console Amazon, choisissez le cluster qui utilise EMRFS CV.
 - b. En haut de la page Détails du cluster, sélectionnez Cloner.
 - c. Choisissez Précédent et passez à l'Étape 1 : Logiciel et étapes.
 - d. Dans Modifier les paramètres du logiciel, supprimez le EMRFS CV. Dans Modifier la configuration, supprimez les configurations suivantes dans la classification `emrfs-site`. Si vous chargez JSON depuis un compartiment S3, vous devez modifier votre objet S3.

```
[
  {"classification":
    "emrfs-site",
    "properties": {
      "fs.s3.consistent.retryPeriodSeconds":"10",
      "fs.s3.consistent":"true",
      "fs.s3.consistent.retryCount":"5",
      "fs.s3.consistent.metadata.tableName":"EmrFSMetadata"
    }
  }
]
```

3. Pour supprimer un EMRFS CV d'un cluster qui utilise des groupes d'instances

- a. Utilisez la commande suivante pour vérifier si un seul EMR cluster utilise la table DynamoDB EMRFS associée à CV ou si plusieurs clusters partagent la table. Le nom de la table est spécifié dans `fs.s3.consistent.metadata.tableName`, comme décrit dans [Configuration de la vue cohérente](#). Le nom de table par défaut utilisé par EMRFS CV est `EmrFSMetadata`.

```
aws emr describe-cluster --cluster-id j-XXXXX | grep
fs.s3.consistent.metadata.tableName
```

- b. Si votre cluster ne partage pas votre base de données DynamoDB avec un autre cluster, utilisez la commande suivante pour reconfigurer le cluster et désactiver CV. EMRFS Pour plus d'informations, consultez [Reconfiguration d'un groupe d'instances dans un cluster en cours d'exécution](#).

```
aws emr modify-instance-groups --cli-input-json file://disable-emrfs-1.json
```

Cette commande ouvre le fichier que vous souhaitez modifier. Modifiez le fichier avec l'une des configurations suivantes.

```
{
  "ClusterId": "j-xxxx",
  "InstanceGroups": [
    {
      "InstanceGroupId": "ig-xxxx",
      "Configurations": [
        {
```

```
"Classification": "emrfs-site",
"Properties": {
  "fs.s3.consistent": "false"
},
"Configurations": []
}
]
}
]
}
```

- c. Si votre cluster partage la table DynamoDB avec un autre cluster, EMRFS désactivez CV sur tous les clusters à un moment où aucun cluster ne modifie les objets de l'emplacement S3 partagé.

Pour supprimer les ressources Amazon DynamoDB associées à un CV EMRFS

Après avoir supprimé le EMRFS CV de vos EMR clusters Amazon, supprimez les ressources DynamoDB associées au CV. EMRFS Jusqu'à ce que vous le fassiez, vous continuez à payer des frais DynamoDB associés à CV. EMRFS

1. Vérifiez les CloudWatch métriques de votre table DynamoDB et confirmez que la table n'est utilisée par aucun cluster.
2. Supprimez la table DynamoDB.

```
aws dynamodb delete-table --table-name <your-table-name>
```

Pour supprimer les SQS ressources Amazon associées à un EMRFS CV

1. Si vous avez configuré votre cluster pour envoyer des notifications d'incohérence à AmazonSQS, vous pouvez supprimer toutes les SQS files d'attente.
2. Trouvez le nom de la SQS file d'attente Amazon spécifié dans `fs.s3.consistent.notification.SQS.queueName`, comme décrit dans [Configurer une vue cohérente](#). Le format de nom de file d'attente par défaut est `EMRFS-Inconsistency-<j-cluster ID>`.

```
aws sqs list-queues | grep 'EMRFS-Inconsistency'
aws sqs delete-queue --queue-url <your-queue-url>
```

Pour arrêter d'utiliser le EMRFS CLI

- [EMRFSCLI](#) gère les métadonnées générées par le EMRFS CV. Comme le support standard pour EMRFS CV arrivera à son terme dans les futures versions d'AmazonEMR, le support pour le EMRFS CLI arrivera également à sa fin.

Rubriques

- [Activation de la vue cohérente](#)
- [Comprendre comment EMRFS une vue cohérente permet de suivre les objets dans Amazon S3](#)
- [Logique des nouvelles tentatives](#)
- [EMRFS métadonnées de vue cohérentes](#)
- [Configurer les notifications de cohérence pour CloudWatch et Amazon SQS](#)
- [Configuration de la vue cohérente](#)
- [EMRFSCLIRéférence de commande](#)

Activation de la vue cohérente

Vous pouvez activer le chiffrement côté serveur Amazon S3 ou la vue cohérente pour EMRFS utiliser le AWS Management Console AWS CLI, ou la classification de `emrfs-site` configuration.

Pour configurer la vue cohérente à l'aide de la console

1. Accédez à la nouvelle EMR console Amazon et sélectionnez **Basculer vers l'ancienne console** dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).
2. Choisissez **Créer un cluster** et **Go to advanced options** (Aller aux options avancées).
3. Choisissez des valeurs pour **Step 1: Software and Steps** (Étape 1 : Logiciel et étapes) et **Step 2: Hardware** (Étape 2 : Matériel).
4. Pour l'étape 3 : Paramètres généraux du cluster, sous **Options supplémentaires**, choisissez **EMRFS une vue cohérente**.
5. Pour le magasin de EMRFS métadonnées, tapez le nom de votre magasin de métadonnées. La valeur par défaut est **EmrFSMetadata**. Si la `mrFSMetadata` table E n'existe pas, elle est créée pour vous dans DynamoDB.

Note

Amazon EMR ne supprime pas automatiquement les EMRFS métadonnées de DynamoDB lorsque le cluster est résilié.

6. Pour Number of retries (Nombre de nouvelles tentatives), tapez une valeur entière. Si une incohérence est détectée, EMRFS essaie d'appeler Amazon S3 autant de fois. La valeur par défaut est **5**.
7. Pour Retry period (in seconds) (Période de nouvelle tentative (en secondes)), tapez une valeur entière. Il s'agit du temps d'EMRFSattente entre deux nouvelles tentatives. La valeur par défaut est **10**.

Note

Les nouvelles tentatives ultérieures utilisent une interruption exponentielle.

Pour lancer un cluster avec une vue cohérente activée à l'aide du AWS CLI

Nous vous recommandons d'installer la version actuelle de AWS CLI. Pour télécharger la dernière version, voir <https://aws.amazon.com/cli/>.

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

```
aws emr create-cluster --instance-type m5.xlarge --instance-count 3 --emrfs  
Consistent=true \  
--release-label emr-7.2.0 --ec2-attributes KeyName=myKey
```

Pour vérifier si la vue cohérente est activée à l'aide du AWS Management Console

- Pour vérifier si la vue cohérente est activée dans la console, accédez à la Liste de clusters et sélectionnez le nom de votre cluster pour afficher Cluster Details (Détails de cluster). Le champ « vue EMRFS cohérente » a une valeur de `Enabled` ou `Disabled`.

Pour vérifier si la vue cohérente est activée en examinant le fichier `emrfs-site.xml`

- Vous pouvez vérifier si la cohérence est activée en inspectant le fichier de configuration `emrfs-site.xml` sur le nœud maître du cluster. Si la valeur booléenne pour `fs.s3.consistent` est définie sur `true`, la vue cohérente est activée pour les opérations du système de fichiers impliquant Amazon S3.

Comprendre comment EMRFS une vue cohérente permet de suivre les objets dans Amazon S3

EMRFS crée une vue cohérente des objets dans Amazon S3 en ajoutant des informations sur ces objets aux EMRFS métadonnées. EMRFS ajoute ces listes à ses métadonnées lorsque :

- Un objet écrit par dans le EMRFS cadre d'une EMR tâche Amazon.
- Un objet est synchronisé avec les EMRFS métadonnées ou importé dans celles-ci à l'aide du EMRFSCLI.

Les objets lus ne EMRFS sont pas automatiquement ajoutés aux métadonnées. Lors de la EMRFS suppression d'un objet, une liste reste dans les métadonnées avec un état supprimé jusqu'à ce que cette liste soit purgée à l'aide du EMRFS CLI. Pour en savoir plus sur le CLI, voir [EMRFSCLI Référence de commande](#). Pour plus d'informations sur la purge des listes dans les EMRFS métadonnées, consultez [EMRFS métadonnées de vue cohérentes](#).

Pour chaque opération Amazon S3, EMRFS vérifie les métadonnées pour obtenir des informations sur l'ensemble d'objets de manière cohérente. S'il EMRFS s'avère qu'Amazon S3 est incohérent au cours de l'une de ces opérations, il tente à nouveau l'opération conformément aux paramètres définis dans les propriétés `emrfs-site` de configuration. Après avoir EMRFS épuisé les tentatives, il lance une exception `ConsistencyException` ou enregistre l'exception et poursuit le flux de travail. Pour plus d'informations sur la logique des nouvelles tentatives, consultez [Logique des nouvelles](#)

[tentatives](#). Vous pouvez trouver des exceptions `ConsistencyExceptions` dans vos journaux, par exemple :

- `listStatus`: aucun objet Amazon S3 pour l'élément de métadonnées `/S3_bucket/dir/object`
- `getFileStatus`: la clé `dir/file` est présente dans les métadonnées, mais pas dans Amazon S3

Si vous supprimez un objet directement d'Amazon S3 suivi par des vues EMRFS cohérentes, considérez EMRFS cet objet comme incohérent car il est toujours répertorié dans les métadonnées telles qu'elles sont présentes dans Amazon S3. Si vos métadonnées ne sont plus synchronisées avec le EMRFS suivi des objets dans Amazon S3, vous pouvez utiliser la `sync` sous-commande du EMRFS CLI pour réinitialiser les métadonnées afin qu'elles reflètent Amazon S3. Pour découvrir les différences entre les métadonnées et Amazon S3, utilisez l'outil `diff`. Enfin, il EMRFS ne dispose que d'une vue cohérente des objets référencés dans les métadonnées ; il se peut que d'autres objets situés sur le même chemin Amazon S3 ne soient pas suivis. Lorsqu'il EMRFS répertorie les objets d'un chemin Amazon S3, il renvoie le surensemble des objets suivis dans les métadonnées et ceux de ce chemin Amazon S3.

Logique des nouvelles tentatives

EMRFS essaie de vérifier la cohérence de la liste pour les objets suivis dans ses métadonnées pendant un certain nombre de tentatives. La valeur par défaut est 5. Au cas où le nombre de nouvelles tentatives est dépassé, la tâche initiale retourne un échec, à moins que `fs.s3.consistent.throwExceptionOnInconsistency` ait la valeur `false`, auquel cas elle consignera uniquement les objets suivis comme incohérents. EMRFS utilise par défaut une politique de réessai exponentielle, mais vous pouvez également la définir sur une politique fixe. Les utilisateurs peuvent également réessayer pendant un certain temps avant de passer au reste de leur tâche sans lever d'exception. Ils peuvent y parvenir en définissant `fs.s3.consistent.throwExceptionOnInconsistency` sur `false`, `fs.s3.consistent.retryPolicyType` sur `fixed` et `fs.s3.consistent.retryPeriodSeconds` sur la valeur de leur choix. L'exemple suivant crée un cluster avec la cohérence activée, qui consigne les incohérences et définit un intervalle fixe de nouvelle tentative de 10 secondes :

Exemple Configuration de la période de nouvelle tentative sur une valeur fixe

```
aws emr create-cluster --release-label emr-7.2.0 \  
--instance-type m5.xlarge --instance-count 1 \  

```

```
--emrfs Consistent=true,Args=[fs.s3.consistent.throwExceptionOnInconsistency=false,  
fs.s3.consistent.retryPolicyType=fixed,fs.s3.consistent.retryPeriodSeconds=10] --ec2-  
attributes KeyName=myKey
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

Pour plus d'informations, consultez [Vue cohérente](#).

EMRFSconfigurations pour IMDS obtenir des appels régionaux

EMRFSs'appuie sur le IMDS (service de métadonnées d'instance) pour obtenir la région de l'instance et Amazon S3, DynamoDB ou les points de terminaison. AWS KMS Toutefois, IMDS le nombre de demandes qu'il peut traiter est limité, et les demandes qui dépassent cette limite échoueront. Cette IMDS limite peut entraîner des EMRFS échecs d'initialisation et entraîner l'échec de la requête ou de la commande. Vous pouvez utiliser le mécanisme de backoff exponentiel aléatoire suivant et les propriétés de configuration d'une région de secours dans le fichier emrfs-site.xml pour résoudre le scénario dans lequel toutes les tentatives échouent.

```
<property>  
  <name>fs.s3.region.retryCount</name>  
  <value>3</value>  
  <description>  
    Maximum retries that would be attempted to get AWS region.  
  </description>  
</property>  
<property>  
  <name>fs.s3.region.retryPeriodSeconds</name>  
  <value>3</value>  
  <description>  
    Base sleep time in second for each get-region retry.  
  </description>  
</property>  
<property>  
  <name>fs.s3.region.fallback</name>  
  <value>us-east-1</value>  
  <description>
```

```
    Fallback to this region after maximum retries for getting AWS region have been
    reached.
  </description>
</property>
```

EMRFS métadonnées de vue cohérentes

EMRFS une vue cohérente assure le suivi de la cohérence à l'aide d'une table DynamoDB pour suivre les objets dans Amazon S3 qui ont été synchronisés avec ou créés par EMRFS. Les métadonnées sont utilisées pour suivre toutes les opérations (lecture, écriture, mise à jour et copie) et aucun contenu réel n'y est stocké. Ces métadonnées sont utilisées pour vérifier si les objets ou les métadonnées provenant d'Amazon S3 correspondent à ce qui est prévu. Cette confirmation permet de vérifier EMRFS la cohérence de la liste et read-after-write la cohérence des nouveaux objets écrits EMRFS sur Amazon S3 ou des objets synchronisés avec EMRFS. Plusieurs clusters peuvent partager les mêmes métadonnées.

Comment ajouter des entrées aux métadonnées

Vous pouvez utiliser les sous-commandes `sync` ou `import` pour ajouter des entrées aux métadonnées. `sync` reflète l'état des objets Amazon S3 dans un chemin d'accès lorsque la sous-commande `import` est utilisée strictement pour ajouter de nouvelles entrées aux métadonnées. Pour plus d'informations, consultez [EMRFS CLI Référence de commande](#).

Vérification des différences entre les métadonnées et les objets dans Amazon S3

Pour vérifier les différences entre les métadonnées et Amazon S3, utilisez la `diff` sous-commande du EMRFS CLI. Pour plus d'informations, consultez [EMRFS CLI Référence de commande](#).

Comment savoir si les opérations de métadonnées sont limitées

EMRFS définit les limites de capacité de débit par défaut des métadonnées pour ses opérations de lecture et d'écriture à 500 et 100 unités, respectivement. Un grand nombre d'objets ou de compartiments peut amener les opérations à dépasser cette capacité, auquel cas DynamoDB limitera les opérations. Par exemple, une application peut provoquer le EMRFS lancement d'un fichier `ProvisionedThroughputExceededException` si vous effectuez une opération qui dépasse ces limites de capacité. Lors de la limitation, l'EMRFS CLI outil tente de réessayer d'écrire dans la table DynamoDB en utilisant un [retard exponentiel jusqu'à la fin de l'opération ou jusqu'à ce qu'il](#) atteigne la valeur maximale de nouvelles tentatives pour écrire des objets d'Amazon vers Amazon S3. EMR

Vous pouvez configurer vos propres limites de capacité de débit. Cependant, DynamoDB impose des limites de partition strictes de 3 000 unités de capacité de lecture RCUs () et de 1 000 unités

de capacité d'écriture WCUs () par seconde pour les opérations de lecture et d'écriture. Pour éviter les sync défaillances causées par la régulation, nous vous recommandons de limiter le débit des opérations de lecture à moins de 3 000 RCUs et celui des opérations d'écriture à moins de 1 000. WCUs Pour obtenir des instructions sur la définition de limites de capacité de débit personnalisées, consultez [Configuration de la vue cohérente](#).

Vous pouvez également consulter CloudWatch les métriques Amazon relatives à vos EMRFS métadonnées dans la console DynamoDB, où vous pouvez voir le nombre de demandes de lecture et d'écriture limitées. Si le nombre de demandes limitées n'est pas nul, votre application peut éventuellement bénéficier d'une capacité grandissante de débit allouée pour les opérations de lecture ou d'écriture. Vous pouvez également réaliser une amélioration des performances si vous voyez que vos opérations approchent de la capacité de débit allouée maximale pour les lectures ou les écritures pendant une période de temps prolongée.

Caractéristiques de débit pour les opérations importantes EMRFS

La valeur par défaut pour les opérations de lecture et d'écriture est de 400 et de 100 unités de capacité de débit, respectivement. Les caractéristiques de performance suivantes vous donnent une idée du débit requis pour certaines opérations. Ces tests ont été effectués à l'aide d'un cluster m3.large à un seul nœud. Toutes les opérations étaient à thread unique. Les performances varient considérablement en fonction des caractéristiques d'application spécifiques et il peut être nécessaire d'expérimenter afin d'optimiser les opérations de système de fichiers.

Opération	Moyenne read-per-second	Moyenne write-per-second
create (objet)	26,79	6,70
delete (objet)	10,79	10,79
delete (répertoire contenant 1 000 objets)	21,79	338,40
getFileStatus(objet)	34,70	0
getFileStatus(annuaire)	19,96	0
listStatus(répertoire contenant 1 objet)	43,31	0

Opération	Moyenne read-per-second	Moyenne write-per-second
listStatus(répertoire contenant 10 objets)	44,34	0
listStatus(répertoire contenant 100 objets)	84,44	0
listStatus(répertoire contenant 1 000 objets)	308,81	0
listStatus(répertoire contenant 10 000 objets)	416,05	0
listStatus(répertoire contenant 100 000 objets)	823,56	0
listStatus(répertoire contenant 1 million d'objets)	882,36	0
mkdir (en continu pendant 120 secondes)	24,18	4,03
mkdir	12,59	0
rename (objet)	19,53	4,88
rename (répertoire contenant 1 000 objets)	23,22	339,34

Pour soumettre une étape qui purge les anciennes données à partir de votre magasin de métadonnées

Les utilisateurs peuvent supprimer des entrées particulières dans les métadonnées basées sur DynamoDB. Cela peut aider à réduire les coûts de stockage associés à la table. Les utilisateurs ont la possibilité de purger des entrées spécifiques manuellement ou par programmation à l'aide de la EMRFS CLI de `le t e` sous-commande. Toutefois, si vous supprimez des entrées des métadonnées, il EMRFS n'est plus possible de vérifier la cohérence.

La purge par programmation après la fin d'une tâche peut être effectuée en soumettant une dernière étape à votre cluster, qui exécute une commande sur le. EMRFS CLI Par exemple, tapez la commande suivante pour soumettre une étape à votre cluster afin de supprimer toutes les entrées de plus de deux jours.

```
aws emr add-steps --cluster-id j-2AL4XXXXXX5T9 --steps Name="emrfsCLI",Jar="command-  
runner.jar",Args=["emrfs","delete","--time","2","--time-unit","days"]  
{  
  "StepIds": [  
    "s-B12345678902"  
  ]  
}
```

Utilisez la StepId valeur renvoyée pour vérifier le résultat de l'opération dans les journaux.

Configurer les notifications de cohérence pour CloudWatch et Amazon SQS

Vous pouvez activer CloudWatch les métriques et les SQS messages Amazon en cas EMRFS de problèmes de cohérence éventuels sur Amazon S3.

CloudWatch

Lorsque CloudWatch les métriques sont activées, une métrique nommée Inconsistency est envoyée chaque fois qu'un FileSystem API appel échoue en raison de la cohérence éventuelle d'Amazon S3.

Pour consulter CloudWatch les statistiques relatives aux éventuels problèmes de cohérence d'Amazon S3

Pour afficher la métrique d'incohérence dans la CloudWatch console, sélectionnez les EMRFS métriques, puis sélectionnez une paire JobFlowId/Metric Name. Par exemple : j-162XXXXXXM2CU ListStatus, j-162XXXXXXM2CU GetFileStatus, etc.

1. Ouvrez la CloudWatch console à l'adresse <https://console.aws.amazon.com/cloudwatch/>.
2. Dans le tableau de bord, dans la section Métriques, choisissez EMRFS.
3. Dans le volet Job Flow Metrics, sélectionnez une ou plusieurs paires JobFlowId/Metric Name. Une représentation graphique des métriques s'affiche dans la fenêtre ci-dessous.

Amazon SQS

Lorsque SQS les notifications Amazon sont activées, une SQS file d'attente Amazon portant le nom `EMRFS-Inconsistency-<jobFlowId>` est créée lors EMRFS de l'initialisation. Les SQS messages Amazon sont placés dans la file d'attente lorsqu'un `FileSystem API` appel échoue en raison de la cohérence éventuelle d'Amazon S3. Le message contient des informations telles qu' `JobFlowId` une liste de chemins incohérents, une trace de pile, etc. API Les messages peuvent être lus à l'aide de la SQS console Amazon ou à l'aide de la EMRFS `read-sqs` commande.

Pour gérer les SQS messages Amazon pour les éventuels problèmes de cohérence d'Amazon S3

SQS Les messages Amazon relatifs aux éventuels problèmes de cohérence d'Amazon S3 peuvent être lus à l'aide du EMRFSCLI. Pour lire les messages d'une SQS file d'attente EMRFS Amazon, tapez la `read-sqs` commande et spécifiez un emplacement de sortie sur le système de fichiers local du nœud principal pour le fichier de sortie obtenu.

Vous pouvez également supprimer une file d'SQSattente EMRFS Amazon à l'aide de la `delete-sqs` commande.

1. Pour lire les messages d'une SQS file d'attente Amazon, tapez la commande suivante. Remplacez *queuename* avec le nom de la SQS file d'attente Amazon que vous avez configurée et remplacée */path/filename* avec le chemin du fichier de sortie :

```
emrfs read-sqs --queue-name queuename --output-file /path/filename
```

Par exemple, pour lire et sortir SQS des messages Amazon depuis la file d'attente par défaut, tapez :

```
emrfs read-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU --output-file /path/filename
```

Note

Vous pouvez également utiliser les raccourcis `-q` et `-o` à la place de `--queue-name` et `--output-file` respectivement.

2. Pour supprimer une SQS file d'attente Amazon, tapez la commande suivante :

```
emrfs delete-sqs --queue-name queuename
```

Par exemple, pour supprimer la file d'attente par défaut, tapez :

```
emrfs delete-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU
```

Note

Vous pouvez également utiliser le raccourci `-q` à la place de `--queue-name`.

Configuration de la vue cohérente

Vous pouvez configurer d'autres paramètres pour la vue cohérente en les fournissant à l'aide des propriétés de configuration pour les propriétés `emrfs-site`. Par exemple, vous pouvez choisir un débit DynamoDB par défaut différent en fournissant les arguments suivants à l'option, en utilisant CLI `--emrfs` la classification de configuration `emrfs-site` (EMR Amazon versions 4.x et ultérieures uniquement) ou une action bootstrap pour configurer le fichier `emrfs-site.xml` sur le nœud principal :

Exemple Modification des valeurs de lecture et d'écriture de métadonnées par défaut au lancement du cluster

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge \  
--emrfs Consistent=true,Args=[fs.s3.consistent.metadata.read.capacity=600,\  
fs.s3.consistent.metadata.write.capacity=300] --ec2-attributes KeyName=myKey
```

Vous pouvez également utiliser le fichier de configuration suivant et l'enregistrer localement ou dans Amazon S3 :

```
[  
  {  
    "Classification": "emrfs-site",  
    "Properties": {  
      "fs.s3.consistent.metadata.read.capacity": "600",  
      "fs.s3.consistent.metadata.write.capacity": "300"  
    }  
  }  
]
```

Utilisez la configuration que vous avez créée avec la syntaxe suivante :

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Hive \
--instance-type m5.xlarge --instance-count 2 --configurations file://./myConfig.json
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

Les options suivantes peuvent être définies à l'aide de configurations ou d' AWS CLI --emrfsarguments. Pour plus d'informations sur ces arguments, consultez la [Référence des commandes AWS CLI](#).

Propriétés **emrfs-site.xml** pour la vue cohérente

Propriété	Valeur par défaut	Description
fs.s3.consistent	false	Lorsqu'elle est définie sur true , cette propriété est configurée EMRFS pour utiliser DynamoDB afin d'assurer la cohérence.
fs.s3.consistent.retryPolicyType	exponential	Cette propriété identifie la stratégie à utiliser lors d'une nouvelle tentative pour des problèmes de cohérence. Les options comprennent : exponential, fixed et none.
fs.s3.consistent.retryPeriodSeconds	1	Cette propriété définit la durée d'attente entre les tentatives de relance de cohérence.
fs.s3.consistent.retryCount	10	Cette propriété définit le nombre maximal de nouvelles tentative

Propriété	Valeur par défaut	Description
		s lorsqu'une incohérence est détectée.
<code>fs.s3.consistent.throwExceptionOnInconsistency</code>	true	Cette propriété détermine s'il convient de lever ou de consigner une exception de cohérence. Lorsque la valeur est true , une exception <code>ConsistencyException</code> est levée.
<code>fs.s3.consistent.metadata.autoCreate</code>	true	Lorsque la valeur est true , cette propriété permet la création automatique de tables de métadonnées.
<code>fs.s3.consistent.metadata.etagVerificationEnabled</code>	true	Avec Amazon EMR 5.29.0, cette propriété est activée par défaut. Lorsque cette option est activée, EMRFS utilise S3 ETags pour vérifier que les objets lus sont de la dernière version disponible. Cette fonctionnalité est utile dans les cas read-after-update d'utilisation dans lesquels des fichiers sur S3 sont remplacés tout en conservant le même nom. Cette fonctionnalité de ETag vérification ne fonctionne pas actuellement avec S3 Select.
<code>fs.s3.consistent.metadata.tableName</code>	EmrFSMetadata	Cette propriété spécifie le nom de la table de métadonnées dans DynamoDB.

Propriété	Valeur par défaut	Description
<code>fs.s3.consistent.metadata.read.capacity</code>	500	Cette propriété spécifie la capacité de lecture DynamoDB à mettre en service lorsque la table de métadonnées est créée.
<code>fs.s3.consistent.metadata.write.capacity</code>	100	Cette propriété spécifie la capacité d'écriture DynamoDB à mettre en service lorsque la table de métadonnées est créée.
<code>fs.s3.consistent.fastList</code>	true	Lorsque la valeur est true , cette propriété utilise plusieurs fils pour répertorier un répertoire (si nécessaire). La cohérence doit être activée pour pouvoir utiliser cette propriété.
<code>fs.s3.consistent.fastList.prefetchMetadata</code>	false	Lorsque la valeur est true , cette propriété permet l'extraction préalable de métadonnées pour les répertoires contenant plus de 20 000 éléments.
<code>fs.s3.consistent.notification.CloudWatch</code>	false	Lorsqu'elle est définie sur true , CloudWatch les métriques sont activées pour les FileSystem API appels qui échouent en raison d'éventuels problèmes de cohérence avec Amazon S3.
<code>fs.s3.consistent.notification.SQS</code>	false	Lorsque ce paramètre est défini sur true , les éventuelles notifications de cohérence sont envoyées vers une SQS file d'attente Amazon.

Propriété	Valeur par défaut	Description
<code>fs.s3.consistent.notification.SQS.queueName</code>	EMRFS-Inconsistency- <jobFlowId>	La modification de cette propriété vous permet de spécifier votre propre nom de SQS file d'attente pour les messages concernant d'éventuels problèmes de cohérence avec Amazon S3.
<code>fs.s3.consistent.notification.SQS.customMsg</code>	none	Cette propriété vous permet de spécifier des informations personnalisées incluses dans les SQS messages concernant d'éventuels problèmes de cohérence sur Amazon S3. Si une valeur n'est pas spécifiée pour cette propriété, le champ correspondant dans le message est vide.
<code>fs.s3.consistent.dynamodb.endpoint</code>	none	Cette propriété vous permet de spécifier un point de terminaison DynamoDB personnalisé pour vos métadonnées de vue cohérente.
<code>fs.s3.useRequesterPaysHeader</code>	false	Lorsqu'elle est définie sur true , cette propriété autorise les requêtes Amazon S3 à être envoyées dans des compartiments avec l'option de payeur de demandes activée.

EMRFSCLIRéférence de commande

EMRFSCLIII est installé par défaut sur tous les nœuds maîtres du cluster créés à l'aide de la EMR version 3.2.1 ou ultérieure d'Amazon. Vous pouvez utiliser le EMRFS CLI pour gérer les métadonnées afin d'obtenir un affichage cohérent.

Note

La emrfs commande n'est prise en charge qu'avec l'émulation du terminal VT1 00. Cependant, elle peut fonctionner avec les autres modes d'émulateur de terminal.

Commande emrfs de niveau supérieur

La commande emrfs de niveau supérieur prend en charge la structure suivante.

```
emrfs [describe-metadata | set-metadata-capacity | delete-metadata | create-metadata |
 \
 list-metadata-stores | diff | delete | sync | import ] [options] [arguments]
```

Spécifiez des [options], avec ou sans arguments, comme décrit dans le tableau suivant. Pour les [options] spécifiques à des sous-commandes (describe-metadata, set-metadata-capacity, etc.), consultez chaque sous-commande ci-dessous.

[Options] pour emrfs

Option	Description	Obligatoire
-a <i>AWS_ACCESS_KEY_ID</i> --access-key <i>AWS_ACCESS_KEY_ID</i>	La clé AWS d'accès que vous utilisez pour écrire des objets sur Amazon S3 et pour créer ou accéder à un magasin de métadonnées dans DynamoDB. Par défaut, <i>AWS_ACCESS_KEY_ID</i> est défini sur la clé d'accès utilisée pour créer le cluster.	Non
	La clé AWS secrète associée à la clé d'accès que vous utilisez pour écrire des objets sur Amazon	Non

Option	Description	Obligatoire
<code>-s <i>AWS_SECRET_ACCESS_KEY</i> --secret-key <i>AWS_SECRET_ACCESS_KEY</i></code>	S3 et pour créer ou accéder à un magasin de métadonnées dans DynamoDB. Par défaut, <i>AWS_SECRET_ACCESS_KEY</i> est défini sur la clé secrète associée à la clé d'accès utilisée pour créer le cluster.	
<code>-v --verbose</code>	Détaillez le résultat.	Non
<code>-h --help</code>	Affiche le message d'aide pour la commande <code>emrfs</code> avec une instruction d'utilisation.	Non

Sous commande `describe-metadata emrfs`

[Options] pour `emrfs describe-metadata`

Option	Description	Obligatoire
<code>-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> est le nom de la table de métadonnées DynamoDB. Si l'icône <i>METADATA_NAME</i> l'argument n'est pas fourni, la valeur par défaut est <code>EmrFSMetadata</code> .	Non

Exemple Exemple `describe-metadata emrfs`

L'exemple suivant décrit la table de métadonnées par défaut.

```
$ emrfs describe-metadata
EmrFSMetadata
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 12
```


sous-commande emrfs set-metadata-capacity

[Options] pour emrfs set-metadata-capacity

Option	Description	Obligatoire
<code>-m <i>METADATA_NAME</i></code> <code>--metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> est le nom de la table de métadonnées DynamoDB. Si l'icône <i>METADATA_NAME</i> l'argument n'est pas fourni, la valeur par défaut est <code>EmrFSMetadata</code> .	Non
<code>-r <i>READ_CAPACITY</i></code> <code>--read-capacity <i>READ_CAPACITY</i></code>	La capacité de débit de lecture demandée pour la table de métadonnées. Si l'icône <i>READ_CAPACITY</i> l'argument n'est pas fourni, la valeur par défaut est <code>400</code> .	Non
<code>-w <i>WRITE_CAPACITY</i></code> <code>--write-capacity <i>WRITE_CAPACITY</i></code>	La capacité de débit d'écriture demandée pour la table de métadonnées. Si l'icône <i>WRITE_CAPACITY</i> l'argument n'est pas fourni, la valeur par défaut est <code>100</code> .	Non

Exemple exemple d'emrfs set-metadata-capacity

L'exemple suivant définit la capacité de débit de lecture sur `600` et la capacité d'écriture sur `150` pour une table de métadonnées nommée `EmrMetadataAlt`.

```
$ emrfs set-metadata-capacity --metadata-name EmrMetadataAlt --read-capacity 600 --
write-capacity 150
  read-capacity: 400
  write-capacity: 100
  status: UPDATING
  approximate-item-count (6 hour delay): 0
```

Sous commande delete-metadata emrfs

[Options] pour emrfs delete-metadata

Option	Description	Obligatoire
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> est le nom de la table de métadonnées DynamoDB. Si l'icône <i>METADATA_NAME</i> l'argument n'est pas fourni, la valeur par défaut est <code>EmrFSMetadata</code> .	Non

Exemple Exemple delete-metadata emrfs

L'exemple suivant supprime la table de métadonnées par défaut.

```
$ emrfs delete-metadata
```

Sous commande create-metadata emrfs

[Options] pour emrfs create-metadata

Option	Description	Obligatoire
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> est le nom de la table de métadonnées DynamoDB. Si l'icône <i>METADATA_NAME</i> l'argument n'est pas fourni, la valeur par défaut est <code>EmrFSMetadata</code> .	Non
-r <i>READ_CAPACITY</i> --read-capacity <i>READ_CAPACITY</i>	La capacité de débit de lecture demandée pour la table de métadonnées. Si l'icône <i>READ_CAPACITY</i> l'argument n'est pas fourni, la valeur par défaut est <code>400</code> .	Non
		Non

Option	Description	Obligatoire
<code>-w <i>WRITE_CAPACITY</i> --write-capacity <i>WRITE_CAPACITY</i></code>	La capacité de débit d'écriture demandée pour la table de métadonnées. Si l'icône <i>WRITE_CAPACITY</i> l'argument n'est pas fourni, la valeur par défaut est 100.	

Exemple Exemple create-metadata emrfs

L'exemple suivant crée une table de métadonnée nommée `EmrFSMetadataAlt`.

```
$ emrfs create-metadata -m EmrFSMetadataAlt
Creating metadata: EmrFSMetadataAlt
EmrFSMetadataAlt
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 0
```

sous-commande emrfs list-metadata-stores

La sous-commande `emrfs list-metadata-stores` n'a aucune [option].

Exemple list-metadata-stores Exemple L

L'exemple suivant répertorie vos tables de métadonnées.

```
$ emrfs list-metadata-stores
EmrFSMetadata
```

Sous-commande diff emrfs

[Options] pour `emrfs diff`

Option	Description	Obligatoire
	<i>METADATA_NAME</i> est le nom de la table de métadonnées DynamoDB. Si l'icône <i>METADATA_</i>	Non

Option	Description	Obligatoire
<code>-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i></code>	<i>NAME</i> l'argument n'est pas fourni, la valeur par défaut est <code>EmrFSMetadata</code> .	
<code>s3://s3Path</code>	Le chemin d'accès au compartiment Amazon S3 à comparer à la table des métadonnées. Synchronisation de compartiments de façon récursive.	Oui

Exemple Exemple diff emrfs

L'exemple suivant compare la table de métadonnées par défaut dans un compartiment Amazon S3.

```
$ emrfs diff s3://elasticmapreduce/samples/cloudfront
BOTH | MANIFEST ONLY | S3 ONLY
DIR elasticmapreduce/samples/cloudfront
DIR elasticmapreduce/samples/cloudfront/code/
DIR elasticmapreduce/samples/cloudfront/input/
DIR elasticmapreduce/samples/cloudfront/logprocessor.jar
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-14.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-15.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-16.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-17.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-18.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-19.WxYz1234
DIR elasticmapreduce/samples/cloudfront/input/XABCD12345678.2009-05-05-20.WxYz1234
DIR elasticmapreduce/samples/cloudfront/code/cloudfront-loganalyzer.tgz
```

Sous commande delete emrfs

[Options] pour emrfs delete

Option	Description	Obligatoire
		Non

Option	Description	Obligatoire
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> est le nom de la table de métadonnées DynamoDB. Si l'icône <i>METADATA_NAME</i> l'argument n'est pas fourni, la valeur par défaut est <code>EmrFSMetadata</code> .	
<i>s3://s3Path</i>	Le chemin d'accès au compartiment Amazon S3 que vous suivez pour une vue cohérente. Synchronisation de compartiments de façon récursive.	Oui
-t <i>TIME</i> --time <i>TIME</i>	L'heure d'expiration (interprétée à l'aide de l'argument d'unité de temps). Toutes les entrées de métadonnées antérieures au <i>TIME</i> les arguments sont supprimés pour le compartiment spécifié.	
-u <i>UNIT</i> --time-unit <i>UNIT</i>	La mesure utilisée pour interpréter l'argument temps (nanosecondes, microsecondes, millisecondes, secondes, minutes, heures ou jours). Si aucun argument n'est spécifié, la valeur par défaut est <code>days</code> .	
--read-consumption <i>READ_CONSUMPTION</i>	Le montant requis de débit de lecture disponible utilisé pour l'opération delete. Si l'icône <i>READ_CONSUMPTION</i> l'argument n'est pas spécifié, la valeur par défaut est <code>400</code> .	Non
--write-consumption <i>WRITE_CONSUMPTION</i>	Le montant requis de débit d'écriture disponible utilisé pour l'opération delete. Si l'icône <i>WRITE_CONSUMPTION</i> l'argument n'est pas spécifié, la valeur par défaut est <code>100</code> .	Non

Exemple Exemple delete emrfs

L'exemple suivant supprime tous les objets dans un compartiment Amazon S3 depuis les métadonnées de suivi pour une vue cohérente.

```
$ emrfs delete s3://elasticmapreduce/samples/cloudfront
entries deleted: 11
```

Sous commande import emrfs

[Options] pour emrfs import

Option	Description	Obligatoire
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> est le nom de la table de métadonnées DynamoDB. Si l'icône <i>METADATA_NAME</i> l'argument n'est pas fourni, la valeur par défaut est <code>EmrFSMetadata</code> .	Non
<i>s3://s3Path</i>	Le chemin d'accès au compartiment Amazon S3 que vous suivez pour une vue cohérente. Synchronisation de compartiments de façon récursive.	Oui
--read-consumption <i>READ_CONSUMPTION</i>	Le montant requis de débit de lecture disponible utilisé pour l'opération delete. Si l'icône <i>READ_CONSUMPTION</i> l'argument n'est pas spécifié, la valeur par défaut est <code>400</code> .	Non
--write-consumption <i>WRITE_CONSUMPTION</i>	Le montant requis de débit d'écriture disponible utilisé pour l'opération delete. Si l'icône <i>WRITE_CONSUMPTION</i> l'argument n'est pas spécifié, la valeur par défaut est <code>100</code> .	Non

Exemple Exemple import emrfs

L'exemple suivant importe tous les objets dans un compartiment Amazon S3 avec les métadonnées de suivi pour une vue cohérente. Toutes les clés inconnus sont ignorées.

```
$ emrfs import s3://elasticmapreduce/samples/cloudfront
```

Sous-commande sync emrfs

[Options] pour emrfs sync

Option	Description	Obligatoire
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> est le nom de la table de métadonnées DynamoDB. Si l'icône <i>METADATA_NAME</i> l'argument n'est pas fourni, la valeur par défaut est <code>EmrFSMetadata</code> .	Non
<i>s3://s3Path</i>	Le chemin d'accès au compartiment Amazon S3 que vous suivez pour une vue cohérente. Synchronisation de compartiments de façon récursive.	Oui
--read-consumption <i>READ_CONSUMPTION</i>	Le montant requis de débit de lecture disponible utilisé pour l'opération delete. Si l'icône <i>READ_CONSUMPTION</i> l'argument n'est pas spécifié, la valeur par défaut est <code>400</code> .	Non
--write-consumption <i>WRITE_CONSUMPTION</i>	Le montant requis de débit d'écriture disponible utilisé pour l'opération delete. Si l'icône <i>WRITE_CONSUMPTION</i> l'argument n'est pas spécifié, la valeur par défaut est <code>100</code> .	Non

Exemple Exemple commande sync emrfs

L'exemple suivant importe tous les objets dans un compartiment Amazon S3 avec les métadonnées de suivi pour une vue cohérente. Toutes les clés inconnues sont supprimées.

```
$ emrfs sync s3://elasticmapreduce/samples/cloudfront
Synching samples/cloudfront                0 added | 0 updated |
 0 removed | 0 unchanged
Synching samples/cloudfront/code/          1 added | 0 updated |
 0 removed | 0 unchanged
Synching samples/cloudfront/               2 added | 0 updated |
 0 removed | 0 unchanged
Synching samples/cloudfront/input/         9 added | 0 updated |
 0 removed | 0 unchanged
Done synching s3://elasticmapreduce/samples/cloudfront 9 added | 0 updated |
 1 removed | 0 unchanged
creating 3 folder key(s)
folders written: 3
```

Sous-commande read-sqs emrfs

[Options] pour emrfs read-sqs

Option	Description	Obligatoire
<code>-q <i>QUEUE_NAME</i> --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> est le nom de la SQS file d'attente Amazon configurée dans <code>emrfs-site.xml</code> . La valeur par défaut est EMRFS-Inconsistency-<i><jobFlowId></i> .	Oui
<code>-o <i>OUTPUT_FILE</i> --output-file <i>OUTPUT_FILE</i></code>	<i>OUTPUT_FILE</i> est le chemin d'accès au fichier de sortie sur le système de fichiers local du nœud maître. Les messages lus depuis la file d'attente sont écrits dans ce fichier.	Oui

Sous commande delete-sqs emrfs

[Options] pour emrfs delete-sqs

Option	Description	Obligatoire
<code>-q <i>QUEUE_NAME</i> --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> est le nom de la SQS file d'attente Amazon configurée dans <code>emrfs-site.xml</code> . La valeur par défaut est EMRFS-Inconsistency-<i><jobFlowId></i> .	Oui

Soumission de EMRFS CLI commandes sous forme d'étapes

L'exemple suivant montre comment utiliser l'`emrfs` utilitaire sur le nœud principal en utilisant le AWS CLI ou API et le `command-runner.jar` pour exécuter la `emrfs` commande en tant qu'étape par étape. L'exemple utilise l'étape AWS SDK for Python (Boto3) pour ajouter une étape à un cluster qui ajoute des objets d'un compartiment Amazon S3 à la table de EMRFS métadonnées par défaut.

```
import boto3
from botocore.exceptions import ClientError

def add_emrfs_step(command, bucket_url, cluster_id, emr_client):
    """
    Add an EMRFS command as a job flow step to an existing cluster.

    :param command: The EMRFS command to run.
    :param bucket_url: The URL of a bucket that contains tracking metadata.
    :param cluster_id: The ID of the cluster to update.
    :param emr_client: The Boto3 Amazon EMR client object.
    :return: The ID of the added job flow step. Status can be tracked by calling
             the emr_client.describe_step() function.
    """
    job_flow_step = {
        "Name": "Example EMRFS Command Step",
        "ActionOnFailure": "CONTINUE",
        "HadoopJarStep": {
            "Jar": "command-runner.jar",
            "Args": ["/usr/bin/emrfs", command, bucket_url],
```

```
    },
}

try:
    response = emr_client.add_job_flow_steps(
        JobFlowId=cluster_id, Steps=[job_flow_step]
    )
    step_id = response["StepIds"][0]
    print(f"Added step {step_id} to cluster {cluster_id}.")
except ClientError:
    print(f"Couldn't add a step to cluster {cluster_id}.")
    raise
else:
    return step_id

def usage_demo():
    emr_client = boto3.client("emr")
    # Assumes the first waiting cluster has EMRFS enabled and has created metadata
    # with the default name of 'EmrFSMetadata'.
    cluster = emr_client.list_clusters(ClusterStates=["WAITING"])["Clusters"][0]
    add_emrfs_step(
        "sync", "s3://elasticmapreduce/samples/cloudfront", cluster["Id"], emr_client
    )

if __name__ == "__main__":
    usage_demo()
```

Vous pouvez utiliser la valeur `step_id` pour vérifier les journaux concernant le résultat de l'opération.

Autoriser l'accès aux EMRFS données dans Amazon S3

Par défaut, le EMR rôle pour EC2 détermine les autorisations d'accès aux EMRFS données dans Amazon S3. Les IAM politiques associées à ce rôle s'appliquent quel que soit l'utilisateur ou le groupe qui fait la demande EMRFS. L'argument par défaut est `EMR_EC2_DefaultRole`. Pour plus d'informations, consultez la section [Rôle de service pour les EC2 instances de cluster \(profil d'EC2instance\)](#).

À partir de la EMR version 5.10.0 d'Amazon, vous pouvez utiliser une configuration de sécurité pour spécifier les IAM rôles pour EMRFS. Cela vous permet de personnaliser les autorisations pour les

EMRFS demandes adressées à Amazon S3 pour les clusters comportant plusieurs utilisateurs. Vous pouvez définir différents IAM rôles pour différents utilisateurs et groupes, ainsi que pour différents emplacements de compartiment Amazon S3 en fonction du préfixe dans Amazon S3. Lorsque vous EMRFS envoyez une demande à Amazon S3 qui correspond à des utilisateurs, à des groupes ou à des emplacements que vous spécifiez, le cluster utilise le rôle correspondant que vous spécifiez au lieu du EMR rôle pour lequel EC2. Pour plus d'informations, consultez [Configurer IAM les rôles pour les EMRFS demandes adressées à Amazon S3](#).

Sinon, si les exigences de votre EMR solution Amazon vont au-delà IAM des rôles EMRFS fournis, vous pouvez définir une classe de fournisseur d'informations d'identification personnalisée, qui vous permet de personnaliser l'accès aux EMRFS données dans Amazon S3.

Création d'un fournisseur d'informations d'identification personnalisé pour EMRFS les données dans Amazon S3

Pour créer un fournisseur d'informations d'identification personnalisé, vous implémentez les [AWSCredentialsProvider](#) classes et Hadoop [Configurable](#).

Pour une explication détaillée de cette approche, voir [Analyser en toute sécurité les données d'un autre AWS compte EMRFS](#) dans le blog AWS Big Data. Le billet de blog inclut un didacticiel qui vous guide tout au long du processus end-to-end, de la création de IAM rôles au lancement du cluster. Il fournit aussi un exemple de code Java qui implémente la classe fournisseur d'informations d'identification personnalisées.

La procédure de base est la suivante :

Pour spécifier un fournisseur d'informations d'identification personnalisées

1. Créez une classe de fournisseur d'informations d'identification personnalisée compilée sous forme de JAR fichier.
2. Exécutez un script sous forme d'action de démarrage pour copier le JAR fichier du fournisseur d'informations d'identification personnalisé à `/usr/share/aws/emr/emrfs/auxlib` emplacement du nœud principal du cluster. Pour plus d'informations sur les actions d'amorçage, consultez [\(Facultatif\) Création d'actions d'amorçage pour installer des logiciels supplémentaires](#).
3. Personnalisez la `emrfs-site` classification pour spécifier la classe que vous implémentez dans le JAR fichier. Pour plus d'informations sur la spécification d'objets de configuration pour

personnaliser les applications, consultez la [section Configuration des applications](#) dans le Amazon EMR Release Guide.

L'exemple suivant illustre une commande `create-cluster` qui lance un cluster Hive avec les paramètres de configuration courants et qui inclut également :

- Une action d'amorçage qui exécute le script, `copy_jar_file.sh`, enregistré sur `mybucket` dans Amazon S3.
- Une `emrfs-site` classification qui spécifie un fournisseur d'informations d'identification personnalisé défini dans le JAR fichier comme `MyCustomCredentialsProvider`

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --applications Name=Hive \
--bootstrap-actions '[{"Path":"s3://mybucket/copy_jar_file.sh","Name":"Custom
action"}]' \
--ec2-attributes '{"KeyName":"MyKeyPair","InstanceProfile":"EMR_EC2_DefaultRole",\
"SubnetId":"subnet-xxxxxxx","EmrManagedSlaveSecurityGroup":"sg-xxxxxxx",\
"EmrManagedMasterSecurityGroup":"sg-xxxxxxx"}' \
--service-role EMR_DefaultRole_V2 --enable-debugging --release-label emr-7.2.0 \
--log-uri 's3n://my-emr-log-bucket/' --name 'test-awscredentialsprovider-emrfs' \
--instance-type=m5.xlarge --instance-count 3 \
--configurations '[{"Classification":"emrfs-site",\
"Properties":
{"fs.s3.customAWSCredentialsProvider":"MyAWSCredentialsProviderWithUri"},\
"Configurations":[]}]'
```

Gestion du point de AWS Security Token Service terminaison par défaut

EMRFS utilise le AWS Security Token Service (STS) pour récupérer des informations d'identification de sécurité temporaires afin d'accéder à vos AWS ressources. Les EMR versions antérieures d'Amazon envoient toutes les AWS STS demandes à un seul point de terminaison mondial à l'adresse `https://sts.amazonaws.com`. Amazon EMR publie les versions 5.31.0 et 6.1.0 et ultérieures pour envoyer des demandes aux points de terminaison régionaux AWS STS. Cela réduit la latence et améliore la validité des jetons de session. Pour plus d'informations sur les AWS STS points de terminaison, consultez [la section Gestion AWS STS dans une AWS région](#) dans le Guide de l'AWS Identity and Access Management utilisateur.

Lorsque vous utilisez les EMR versions 5.31.0 et 6.1.0 d'Amazon et versions ultérieures, vous pouvez remplacer le point de terminaison par défaut. AWS STS Pour ce faire, vous devez modifier la propriété `fs.s3.sts.endpoint` dans votre configuration `emrfs-site`.

L' AWS CLI exemple suivant définit le point de AWS STS terminaison par défaut utilisé par EMRFS le point de terminaison global.

```
aws emr create-cluster --release-label <emr-5.33.0> --instance-type m5.xlarge \  
--emrfs Args=[fs.s3.sts.endpoint=https://sts.amazonaws.com]
```

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

Vous pouvez également créer un fichier de JSON configuration à l'aide de l'exemple suivant et le spécifier à l'aide de l'argument `--configurations` de `emr create-cluster`. Pour plus d'informations sur l'utilisation de `--configurations`, consultez la [Référence des commandes AWS CLI](#).

```
[  
  {  
    "classification": "emrfs-site",  
    "properties": {
```

```
    "fs.s3.sts.endpoint": "https://sts.amazonaws.com"  
  }  
}  
]
```

Spécification du chiffrement Amazon S3 à l'aide EMRFS des propriétés

Important

À partir de la EMR version 4.8.0 d'Amazon, vous pouvez utiliser les configurations de sécurité pour appliquer les paramètres de chiffrement plus facilement et avec davantage d'options. Nous vous recommandons d'utiliser les configurations de sécurité. Pour plus d'informations, consultez [Configuration du chiffrement des données](#). Les instructions de console décrites dans cette section sont disponibles pour les versions antérieures à la version 4.8.0. Si vous utilisez le AWS CLI pour configurer le chiffrement Amazon S3 à la fois dans la configuration du cluster et dans une configuration de sécurité dans les versions suivantes, la configuration de sécurité remplace la configuration du cluster.


Lorsque vous créez un cluster, vous pouvez spécifier le chiffrement côté serveur (SSE) ou le chiffrement côté client (CSE) pour les données d'EMRFS Amazon S3 à l'aide de la console ou à l'aide des propriétés de `emrfs-site classification` via le ou. AWS CLI EMR SDK Amazon S3 SSE et Amazon S3 s'CSE excluent mutuellement ; vous pouvez choisir l'un ou l'autre, mais pas les deux.

Pour AWS CLI obtenir des instructions, consultez la section correspondant à votre type de chiffrement ci-dessous.

Pour définir les options EMRFS de chiffrement à l'aide du AWS Management Console

1. Accédez à la nouvelle EMR console Amazon et sélectionnez **Basculer vers l'ancienne console** dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).
2. Choisissez **Créer un cluster** et **Go to advanced options** (Aller aux options avancées).
3. Choisissez une **Release (Version) 4.7.2** ou antérieure.
4. Choisissez d'autres options pour **Software and Steps** (Logiciel et étapes) selon les besoins de votre application, puis cliquez sur **Next** (Suivant).

5. Choisissez les paramètres dans les volets Hardware (Matériel) et General Cluster Settings (Paramètres de cluster généraux) selon les besoins de votre application.
6. Dans le volet Sécurité, sous Authentification et chiffrement, sélectionnez l'option Chiffrement S3 (avecEMRFS) à utiliser.

 Note

Le chiffrement S3 côté serveur avec gestion des KMS clés (SSE-KMS) n'est pas disponible lors de l'utilisation de la EMR version 4.4 ou antérieure d'Amazon.

- Si vous choisissez une option qui utilise la gestion des AWS clés, choisissez un ID de AWS KMS clé. Pour plus d'informations, consultez [Utilisation AWS KMS keys pour le EMRFS chiffrement](#).
 - Si vous choisissez le chiffrement côté client S3 avec un fournisseur de matériel personnalisé, indiquez le nom de la classe et l'JARemplacement. Pour plus d'informations, consultez [Chiffrement côté client sur Amazon S3](#).
7. Choisissez d'autres options selon les besoins de votre application, puis choisissez Create Cluster (Créer un cluster).

Utilisation AWS KMS keys pour le EMRFS chiffrement

La clé de AWS KMS chiffrement doit être créée dans la même région que votre instance de EMR cluster Amazon et les compartiments Amazon S3 utilisés avecEMRFS. Si la clé que vous spécifiez se trouve dans un compte différent de celui que vous utilisez pour configurer un cluster, vous devez spécifier la clé à l'aide de sonARN.

Le rôle du profil d'EC2instance Amazon doit être autorisé à utiliser la KMS clé que vous spécifiez. Le rôle par défaut du profil d'instance dans Amazon EMR estEMR_EC2_DefaultRole. Si vous utilisez un rôle différent pour le profil d'instance, ou si vous utilisez IAM des rôles pour les EMRFS demandes adressées à Amazon S3, assurez-vous que chaque rôle est ajouté en tant qu'utilisateur clé, le cas échéant. Cela donne au rôle l'autorisation d'utiliser la KMS clé. Pour plus d'informations, consultez les [sections Utilisation des politiques clés](#) dans le guide du AWS Key Management Service développeur et [Configuration IAM des rôles pour les EMRFS demandes adressées à Amazon S3](#).

Vous pouvez utiliser le AWS Management Console pour ajouter votre profil d'EC2instance ou votre profil d'instance à la liste des utilisateurs clés pour la KMS clé spécifiée, ou vous pouvez utiliser le AWS CLI ou AWS SDK pour attacher une politique de clé appropriée.

Notez qu'Amazon ne EMR prend en charge que les [KMSclés symétriques](#). Vous ne pouvez pas utiliser de [KMSClé asymétrique](#) pour chiffrer des données au repos dans un cluster AmazonEMR. Pour savoir si une KMS clé est symétrique ou asymétrique, consultez la section [Identification des clés symétriques et asymétriques](#). KMS

La procédure ci-dessous décrit comment ajouter le profil d'EMRinstance Amazon par défaut, `EMR_EC2_DefaultRole` en tant qu'utilisateur clé à l'aide du AWS Management Console. Cela suppose que vous avez déjà créé une KMS clé. Pour créer une nouvelle KMS clé, consultez la section [Création de clés](#) dans le guide du AWS Key Management Service développeur.

Pour ajouter le profil d'EC2instance d'Amazon EMR à la liste des utilisateurs de clés de chiffrement

1. Connectez-vous à la console AWS Key Management Service (AWS KMS) AWS Management Console et ouvrez-la à l'[adresse https://console.aws.amazon.com/kms](https://console.aws.amazon.com/kms).
2. Pour modifier le Région AWS, utilisez le sélecteur de région dans le coin supérieur droit de la page.
3. Sélectionnez l'alias de la KMS clé à modifier.
4. Sur la page de détails de la clé, sous Key Users (Utilisateurs de clés), choisissez Add (Ajouter).
5. Dans la boîte de dialogue Ajouter des utilisateurs clés sélectionnez le rôle approprié. Le nom du rôle par défaut est `EMR_EC2_DefaultRole`.
6. Choisissez Ajouter.

Chiffrement côté serveur sur Amazon S3

Lorsque vous configurez le chiffrement côté serveur sur Amazon S3, Amazon S3 chiffre les données au niveau de l'objet au moment où elles sont écrites sur le disque et déchiffre les données lorsqu'elles sont accédées. Pour plus d'informationsSSE, consultez la section [Protection des données à l'aide du chiffrement côté serveur dans le](#) guide de l'utilisateur d'Amazon Simple Storage Service.

Vous pouvez choisir entre deux systèmes de gestion des clés différents lorsque vous le spécifiez SSE dans Amazon EMR :

- SSE-S3 — Amazon S3 gère les clés pour vous.

- SSE- KMS — Vous utilisez un AWS KMS key pour configurer des politiques adaptées à AmazonEMR. Pour plus d'informations sur les exigences clés pour AmazonEMR, consultez la section [Utilisation à AWS KMS keys des fins de chiffrement](#).

SSE avec les clés fournies par le client (SSE-C) n'est pas disponible pour une utilisation avec Amazon. EMR

Pour créer un cluster avec SSE -S3 activé à l'aide du AWS CLI

- Saisissez la commande suivante :

```
aws emr create-cluster --release-label emr-4.7.2 or earlier \  
--instance-count 3 --instance-type m5.xlarge --emrfs Encryption=ServerSide
```

Vous pouvez également activer SSE -S3 en définissant le `fs.s3.enableServerSide` propriété de chiffrement prend la valeur `true` dans les `emrfs-site` propriétés. Consultez l'exemple de SSE - KMS ci-dessous et omettez la propriété pour Key ID.

Pour créer un cluster avec SSE - KMS activé à l'aide du AWS CLI

Note

SSE- n'KMS est disponible que dans les EMR versions 4.5.0 et ultérieures d'Amazon.

- Tapez la AWS CLI commande suivante pour créer un cluster avec SSE -KMS, où *keyID* est AWS KMS key, par exemple, *a4567b8-9900-12ab-1234-123a45678901*:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier --instance-count 3 \  
--instance-type m5.xlarge --use-default-roles \  
--emrfs Encryption=ServerSide,Args=[fs.s3.serverSideEncryption.kms.keyId=keyId]
```

--OR--

Tapez la AWS CLI commande suivante en utilisant la `emrfs-site` classification et fournissez un JSON fichier de configuration dont le contenu est similaire à celui illustré `myConfig.json` dans l'exemple ci-dessous :

```
aws emr create-cluster --release-label emr-4.7.2 or earlier --instance-count 3
--instance-type m5.xlarge --applications Name=Hadoop --configurations file://
myConfig.json --use-default-roles
```

Exemple de contenu de myConfig.json :

```
[
  {
    "Classification":"emrfs-site",
    "Properties": {
      "fs.s3.enableServerSideEncryption": "true",
      "fs.s3.serverSideEncryption.kms.keyId":"a4567b8-9900-12ab-1234-123a45678901"
    }
  }
]
```

Propriétés de configuration pour SSE -S3 et SSE - KMS

Ces propriétés peuvent être configurées à l'aide de la classification de configuration `emrfs-site`. SSE- n'KMS est disponible que dans les EMR versions 4.5.0 et ultérieures d'Amazon.

Propriété	Valeur par défaut	Description
<code>fs.s3.enableServerSideEncryption</code>	false	Lorsque la valeur est true , les objets stockés dans Amazon S3 sont chiffrés à l'aide du chiffrement côté serveur. Si aucune clé n'est spécifiée, SSE -S3 est utilisé.
<code>fs.s3.serverSideEncryption.kms.keyId</code>	n/a	Spécifie un identifiant de AWS KMS clé ou ARN. Si une clé est spécifiée, SSE - KMS est utilisé.

Chiffrement côté client sur Amazon S3

Avec le chiffrement côté client Amazon S3, le chiffrement et le déchiffrement Amazon S3 ont lieu dans le EMRFS client de votre cluster. Les objets sont chiffrés avant d'être chargés sur Amazon S3 et déchiffrés après leur chargement. Le fournisseur que vous indiquez fournit la clé de chiffrement utilisée par le client. Le client peut utiliser les clés fournies par AWS KMS (CSE-KMS) ou une classe Java personnalisée qui fournit la clé racine côté client (CSE-C). Les spécificités du chiffrement sont légèrement différentes entre CSE - KMS et CSE -C, selon le fournisseur spécifié et les métadonnées de l'objet déchiffré ou chiffré. Pour plus d'informations sur ces différences, consultez [Protection des données à l'aide du chiffrement côté client](#) dans le Guide de l'utilisateur Amazon Simple Storage Service.

Note

Amazon S3 garantit CSE uniquement que les EMRFS données échangées avec Amazon S3 sont chiffrées ; toutes les données présentes sur les volumes d'instance de cluster ne sont pas cryptées. De plus, comme Hue n'en a pas EMRFS besoin, les objets que le navigateur de fichiers Hue S3 écrit sur Amazon S3 ne sont pas chiffrés.

Pour spécifier CSE, KMS pour les EMRFS données d'Amazon S3 à l'aide du AWS CLI

- Tapez la commande suivante et remplacez *MyKMSKeyID* avec l'ID de clé ou ARN de la KMS clé à utiliser :

```
aws emr create-cluster --release-label emr-4.7.2 or earlier
--emrfs Encryption=ClientSide,ProviderType=KMS,KMSKeyId=MyKMSKeyId
```

Création d'un fournisseur de clés personnalisé

Selon le type de chiffrement que vous utilisez lors de la création d'un fournisseur de clés personnalisé, l'application doit également implémenter différentes EncryptionMaterialsProvider interfaces. Les deux interfaces sont disponibles dans les versions 1.11.0 et ultérieures AWS SDK pour Java.

- Pour implémenter le chiffrement Amazon S3, utilisez le fichier [com.amazonaws.services.s3.model.EncryptionMaterialsProvider](#) interface.

- Pour implémenter le chiffrement du disque local, utilisez le fichier [com.amazonaws.services.elasticmapreduce.spi.security.EncryptionMaterialsProvider](https://docs.aws.amazon.com/elasticmapreduce/spi/security/EncryptionMaterialsProvider) interface.

Vous pouvez utiliser n'importe quelle stratégie pour fournir du matériel de chiffrement pour la mise en œuvre. Par exemple, vous pouvez choisir de fournir du matériel de chiffrement statique ou de l'intégrer à un système de gestion de clés plus complexe.

Si vous utilisez le chiffrement Amazon S3, vous devez utiliser les algorithmes de chiffrement AES/GCM/NoPadding pour le matériel de chiffrement personnalisé.

Si vous utilisez le chiffrement de disque local, l'algorithme de chiffrement à utiliser pour le matériel de chiffrement personnalisé varie selon les EMR versions. Pour Amazon EMR 7.0.0 et versions antérieures, vous devez utiliser AES/GCM/NoPadding. Pour Amazon EMR 7.1.0 et versions ultérieures, vous devez utiliser AES.

La `EncryptionMaterialsProvider` classe obtient le matériel de chiffrement par contexte de chiffrement. Amazon EMR renseigne les informations contextuelles de chiffrement au moment de l'exécution pour aider l'appelant à déterminer le matériel de chiffrement à renvoyer.

Exemple Exemple : utilisation d'un fournisseur de clés personnalisé pour le chiffrement Amazon S3 avec EMRFS

Lorsqu'Amazon EMR extrait le matériel de chiffrement de la `EncryptionMaterialsProvider` classe pour effectuer le chiffrement, il remplit EMRFS éventuellement l' `materialsDescription` argument avec deux champs : l'Amazon S3 URI pour l'objet et le champ `JobFlowId` du cluster, qui peuvent être utilisés par la `EncryptionMaterialsProvider` classe pour renvoyer du matériel de chiffrement de manière sélective.

Par exemple, le fournisseur peut renvoyer des clés différentes pour différents URI préfixes Amazon S3. Il s'agit de la description du matériel de chiffrement renvoyé qui est finalement stocké avec l'objet Amazon S3 plutôt que de la `materialsDescription` valeur générée EMRFS et transmise au fournisseur. Lors du déchiffrement d'un objet Amazon S3, la description du matériel de chiffrement est transmise à la `EncryptionMaterialsProvider` classe, afin qu'elle puisse, à nouveau, renvoyer de manière sélective la clé correspondante pour déchiffrer l'objet.

Une implémentation `EncryptionMaterialsProvider` de référence est fournie ci-dessous. Un autre fournisseur personnalisé est disponible auprès de GitHub. [EMRFSRSAEncryptionMaterialsProvider](#)

```
import com.amazonaws.services.s3.model.EncryptionMaterials;
import com.amazonaws.services.s3.model.EncryptionMaterialsProvider;
import com.amazonaws.services.s3.model.KMSEncryptionMaterials;
```

```
import org.apache.hadoop.conf.Configurable;
import org.apache.hadoop.conf.Configuration;

import java.util.Map;

/**
 * Provides KMSEncryptionMaterials according to Configuration
 */
public class MyEncryptionMaterialsProviders implements EncryptionMaterialsProvider,
    Configurable{
    private Configuration conf;
    private String kmsKeyId;
    private EncryptionMaterials encryptionMaterials;

    private void init() {
        this.kmsKeyId = conf.get("my.kms.key.id");
        this.encryptionMaterials = new KMSEncryptionMaterials(kmsKeyId);
    }

    @Override
    public void setConf(Configuration conf) {
        this.conf = conf;
        init();
    }

    @Override
    public Configuration getConf() {
        return this.conf;
    }

    @Override
    public void refresh() {

    }

    @Override
    public EncryptionMaterials getEncryptionMaterials(Map<String, String>
materialsDescription) {
        return this.encryptionMaterials;
    }

    @Override
    public EncryptionMaterials getEncryptionMaterials() {
        return this.encryptionMaterials;
    }
}
```

```
}  
}
```

Spécifier un fournisseur de matériaux personnalisés à l'aide du AWS CLI

Pour utiliser l' AWS CLI, transmettez les arguments `Encryption`, `ProviderType`, `CustomProviderClass` et `CustomProviderLocation` à l'option `emrfs`.

```
aws emr create-cluster --instance-type m5.xlarge --release-label emr-4.7.2 or earlier  
--emrfs Encryption=ClientSide,ProviderType=Custom,CustomProviderLocation=s3://  
mybucket/myfolder/provider.jar,CustomProviderClass=classname
```

Le paramètre `Encryption ClientSide` active le chiffrement côté client. `CustomProviderClass` Il s'agit du nom de votre `EncryptionMaterialsProvider` objet et `CustomProviderLocation` de l'emplacement local ou Amazon S3 à partir duquel Amazon EMR copie les données `CustomProviderClass` vers chaque nœud du cluster et les place dans le chemin de classe.

Spécifier un fournisseur de matériaux personnalisés à l'aide d'un SDK

Pour utiliser un SDK, vous pouvez définir la propriété `fs.s3.cse.encryptedMaterialsProvider.uri` pour télécharger la `EncryptionMaterialsProvider` classe personnalisée que vous stockez dans Amazon S3 sur chaque nœud de votre cluster. Vous le configurez dans un `emrfs-site.xml` fichier avec l'emplacement CSE activé et approprié du fournisseur personnalisé.

Par exemple, lors de l' AWS SDK for Java utilisation `RunJobFlowRequest`, votre code peut ressembler à ce qui suit :

```
<snip>  
Map<String,String> emrfsProperties = new HashMap<String,String>();  
    emrfsProperties.put("fs.s3.cse.encryptedMaterialsProvider.uri","s3://mybucket/  
MyCustomEncryptionMaterialsProvider.jar");  
    emrfsProperties.put("fs.s3.cse.enabled","true");  
    emrfsProperties.put("fs.s3.consistent","true");  
  
emrfsProperties.put("fs.s3.cse.encryptedMaterialsProvider","full.class.name.of.EncryptionMate  
  
Configuration myEmrfsConfig = new Configuration()  
    .withClassification("emrfs-site")  
    .withProperties(emrfsProperties);
```

```
RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("Custom EncryptionMaterialsProvider")
    .withReleaseLabel("emr-7.2.0")
    .withApplications(myApp)
    .withConfigurations(myEmrfsConfig)
    .withServiceRole("EMR_DefaultRole_V2")
    .withJobFlowRole("EMR_EC2_DefaultRole")
    .withLogUri("s3://myLogUri/")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myEc2Key")
        .withInstanceCount(2)
        .withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m5.xlarge")
        .withSlaveInstanceType("m5.xlarge")
    );

RunJobFlowResult result = emr.runJobFlow(request);
</snip>
```

Personnalisé EncryptionMaterialsProvider avec arguments

Vous devrez peut-être transmettre des arguments directement au fournisseur. Pour ce faire, vous pouvez utiliser la classification de configuration `emrfs-site` avec les propriétés définies comme arguments personnalisés. Un exemple de configuration est indiqué ci-dessous, qui est enregistré dans un fichier, `myConfig.json` :

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "myProvider.arg1": "value1",
      "myProvider.arg2": "value2"
    }
  }
]
```

À l'aide de la `create-cluster` commande du AWS CLI, vous pouvez utiliser l'`--configurationoption` pour spécifier le fichier comme indiqué ci-dessous :

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge
--instance-count 2 --configurations file://myConfig.json --emrfs
```

```
Encryption=ClientSide,CustomProviderLocation=s3://mybucket/myfolder/
myprovider.jar,CustomProviderClass=classname
```

Configuration de la prise EMRFS en charge du S3EC V2

SDK Les versions Java S3 (1.11.837 et ultérieures) prennent en charge la version 2 du client de chiffrement (S3EC V2) avec diverses améliorations de sécurité. Pour plus d'informations, consultez l'article de blog S3 [Updates to the Amazon S3 encryption client](#). Reportez-vous également à la [migration du client de chiffrement Amazon S3](#) dans le manuel du AWS SDK for Java développeur.

Le client de chiffrement V1 est toujours disponible dans le SDK pour des raisons de rétrocompatibilité. Par défaut EMRFS, S3EC V1 est utilisé pour chiffrer et déchiffrer les objets S3 s'il est activé. CSE

Les objets S3 chiffrés avec S3EC V2 ne peuvent pas être déchiffrés EMRFS sur un EMR cluster dont la version de publication est antérieure à emr-5.31.0 (emr-5.30.1 et versions antérieures, emr-6.1.0 et versions antérieures).

Exemple Configurer EMRFS pour utiliser S3EC V2

Pour configurer EMRFS afin d'utiliser S3EC V2, ajoutez la configuration suivante :

```
{
  "Classification": "emrfs-site",
  "Properties": {
    "fs.s3.cse.encryptionV2.enabled": "true"
  }
}
```

Propriétés **emrfs-site.xml** pour le chiffrement côté client Amazon S3

Propriété	Valeur par défaut	Description
<code>fs.s3.cse.enabled</code>	false	Lorsqu'il est défini sur true , EMRFS les objets stockés dans Amazon S3 sont chiffrés à l'aide du chiffrement côté client.

Propriété	Valeur par défaut	Description
<code>fs.s3.cse.encryptionV2.enabled</code>	false	Lorsqu'il est défini sur <code>true</code> , EMRFS utilise le client de chiffrement S3 version 2 pour chiffrer et déchiffrer les objets sur S3. Disponible pour les EMR versions 5.31.0 et ultérieures.
<code>fs.s3.cse.encryptionMaterialsProvider.uri</code>	N/A	S'applique lors de l'utilisation de supports de chiffrement personnalisés. L'Amazon S3 URI où se trouve <code>EncryptionMaterialsProvider</code> trouve le JAR <code>with</code> . Lorsque vous le fournissez <code>zURI</code> , Amazon le télécharge EMR automatiquement JAR sur tous les nœuds du cluster.
<code>fs.s3.cse.encryptionMaterialsProvider</code>	N/A	Le chemin de classe <code>EncryptionMaterialsProvider</code> utilisé avec le chiffrement côté client. Lorsque vous utilisez CSE - KMS, spécifiez <code>com.amazonaws.emr.hadoop.fs.cse.KMSEncryptionMaterialsProvider</code> .

Propriété	Valeur par défaut	Description
<code>fs.s3.cse.materialsDescription.enabled</code>	false	Lorsqu'il est défini sur <code>true</code> , remplit le fichier <code>materialsDescription</code> des objets chiffrés avec l'Amazon S3 URI pour l'objet et le <code>JobFlowId</code> . Définissez sur <code>true</code> lors de l'utilisation de supports de chiffrement personnalisés.
<code>fs.s3.cse.kms.keyId</code>	N/A	S'applique lors de l'utilisation de CSE -KMS. La valeur de <code>KeyIdARN</code> , ou alias de la KMS clé utilisée pour le chiffrement.
<code>fs.s3.cse.cryptoStorageMode</code>	ObjectMetadata	Le mode de stockage Amazon S3. Par défaut, la description des informations de chiffrement est stockée dans les métadonnées de l'objet. Vous pouvez également stocker la description dans un fichier d'instructions. Les valeurs valides sont <code>ObjectMetadata</code> et <code>InstructionFile</code> . Pour plus d'informations, consultez la section Chiffrement des données côté client avec Amazon S3 AWS SDK for Java et Amazon S3 .

CloudWatch Agent Amazon

L' CloudWatch agent Amazon sur Amazon EMR est un outil qui permet de surveiller les EC2 instances Amazon de votre EMR cluster. Vous pouvez stocker et consulter les métriques que vous collectez avec l' CloudWatch agent dans CloudWatch. Pour plus d'informations sur l' CloudWatch agent, consultez le [guide de CloudWatch l'utilisateur Amazon](#).

Note

Des frais supplémentaires vous seront facturés si vous en utilisez un autre AWS services pour publier, consulter ou consulter les statistiques des CloudWatch agents Amazon. Consultez les pages suivantes pour plus d'informations sur les tarifs.

- [CloudWatch Tarifs Amazon](#)
- [Tarification du service géré par Amazon pour Prometheus](#)
- [Tarification d'Amazon Managed Grafana](#)

Ces métriques sont distinctes des métriques gratuites qu'Amazon EMR fournit sous l'espace de AWS/ElasticMapReduce noms dans CloudWatch. Pour plus d'informations sur les métriques que l' CloudWatch agent ne publie pas, consultez [Surveiller EMR les métriques Amazon avec CloudWatch](#).

Avec Amazon EMR 7.0 et versions ultérieures, vous pouvez installer une version personnalisée de l' CloudWatch agent Amazon sur votre cluster afin de collecter des métriques à partir de EMR celui-ci. L'agent collecte les métriques de tous les nœuds du cluster, les rassemble sur le nœud principal et publie les métriques de ce nœud dans le cloud.

Vous pouvez installer l'agent lorsque vous créez un nouveau cluster dans la console ou lorsque vous utilisez le `create-cluster` API. Pour plus d'informations, consultez [Créez un EMR cluster qui utilise l' CloudWatch agent Amazon](#). Par défaut, un EMR cluster Amazon qui exécute la série version 7.x publie [34 mesures CloudWatch au niveau du système à des intervalles de 60 secondes, mais vous pouvez configurer l'agent pour qu'il publie différentes métriques](#). Une autre option consiste à publier les statistiques sur Amazon Managed Service for Prometheus, et vous pouvez également choisir les mesures à publier. Pour différents cas d'utilisation et configurations, vous pouvez configurer l'agent pour afficher et interroger les métriques dans la CloudWatch console, Amazon Managed Grafana,

ou via APIs le for CloudWatch ou Amazon Managed Service pour Prometheus. Ils AWS services entraînent des frais lorsque vous les utilisez pour stocker et interroger des métriques.

L' CloudWatch agent sur Amazon EMR peut publier des métriques du système ainsi que JMX des métriques pour les services suivants sur votre EMR cluster Amazon.

- Hadoop DataNode
- Hadoop NameNode
- Laine NodeManager
- Laine ResourceManager
- HBaseMaster — Amazon EMR 7.1 et versions ultérieures uniquement
- HBase RegionServer — Amazon EMR 7.1 et versions ultérieures uniquement
- HBase ThriftServer — Amazon EMR 7.1 et versions ultérieures uniquement

Pour plus d'informations sur les métriques disponibles et sur la façon de configurer l' CloudWatch agent sur AmazonEMR, consultez [Configuration de CloudWatch l'agent pour Amazon EMR](#).

Le tableau suivant répertorie la version AmazonCloudWatchAgent incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants avec lesquels Amazon EMR installe AmazonCloudWatchAgent

Pour la version des composants installés AmazonCloudWatchAgent dans cette version, voir [Versions des composants de la version 7.2.0](#).

AmazonCloudWatchAgent informations de version pour emr-7.2.0

Étiquette EMR de sortie Amazon	AmazonCloudWatchAgent Version	Composants installés avec AmazonCloudWatchAgent
emr-7.2.0	AmazonCloudWatchAgent 1.300032.2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-

Étiquette EMR de sortie Amazon	AmazonCloudWatchAgent Version	Composants installés avec AmazonCloudWatchAgent
		mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Rubriques

- [Créez un EMR cluster qui utilise l' CloudWatch agent Amazon](#)
- [Métriques par défaut pour les CloudWatch agents travaillant avec Amazon EMR](#)
- [Configuration de CloudWatch l'agent pour Amazon EMR](#)
- [Considérations et restrictions](#)
- [CloudWatch historique des versions de l'agent](#)

Créez un EMR cluster qui utilise l' CloudWatch agent Amazon

Les procédures décrites dans cette section décrivent les étapes à suivre pour créer un cluster dans Amazon EMR avec l' CloudWatch agent Amazon à partir du AWS Management Console et du AWS CLI.

Rubriques

- [IAMAutorisations requises pour CloudWatch l'agent](#)
- [Point de terminaison CloudWatch d'agent requis](#)
- [Créer un cluster EMR](#)

IAMAutorisations requises pour CloudWatch l'agent

L' CloudWatch agent a besoin de l'cloudwatch:PutMetricDataautorisation AWS Identity and Access Management (IAM) dans le profil d'EC2instance Amazon pour AmazonEMR. Le rôle EMR par défaut d'Amazon dispose déjà de cette autorisation. Vous pouvez créer le rôle par défaut à partir du AWS CLI withaws emr create-default-roles. Pour plus d'informations, consultez la section [Rôle de service pour les EC2 instances de cluster \(profil d'EC2instance\)](#) dans le guide EMR de gestion Amazon.

L'exemple de IAM politique suivant inclut l'`cloudwatch:PutMetricData` autorisation :

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "cloudwatch:PutMetricData",
      "Resource": "*"
    }
  ]
}
```

Point de terminaison CloudWatch d'agent requis

Pour publier des métriques CloudWatch pour un EMR cluster dans un sous-réseau privé, créez un point de terminaison d'agent CloudWatch et associez-le au point de terminaison dans VPC lequel se trouve le sous-réseau privé.

Pour plus d'informations sur les CloudWatch points de terminaison de chaque Région AWS, consultez la section [CloudWatch Points de terminaison et quotas Amazon](#) dans le Guide de référence AWS général.

Créer un cluster EMR

Une fois que vous avez configuré les autorisations et le point de terminaison requis pour une utilisation avec l'agent CloudWatch, utilisez le AWS Management Console ou le AWS CLI pour créer un nouveau cluster avec l'agent installé.

Console

Pour créer un cluster avec CloudWatch l'agent Amazon depuis la console

1. Accédez à la EMR console Amazon.
2. Choisissez Créer un cluster.
3. Sous Nom et applications, choisissez une EMR version Amazon 7.0.0 ou supérieure.
4. Sous Ensemble d'applications, sélectionnez le ou les applications que vous souhaitez installer sur votre cluster et incluez CloudWatch l'agent dans vos sélections.
5. Procédez à la création du cluster adapté à votre cas d'utilisation.

AWS CLI

Dans le AWS CLI, vous pouvez ajouter CloudWatch l'agent Amazon à un cluster avec le `--applications` paramètre pour `create-cluster`.

Pour créer un cluster avec l' CloudWatch agent Amazon à partir du AWS CLI

- Lorsque vous créez un cluster, utilisez une commande similaire à la suivante pour inclure l' CloudWatch agent Amazon. *myKey* Remplacez-le par le nom de votre paire de EC2 clés.

```
aws emr create-cluster --name "Spark cluster with CloudWatch agent" \  
--release-label emr-7.0.0 \  
--applications Name=Spark Name=AmazonCloudWatchAgent \  
--ec2-attributes KeyName=myKey --instance-type m7g.2xlarge \  
--instance-count 3 --use-default-roles
```

Pour plus de détails sur l'utilisation d'Amazon EMR avec le AWS CLI, consultez la [référence des AWS CLI commandes](#).

Métriques par défaut pour les CloudWatch agents travaillant avec Amazon EMR

Lorsque vous installez l' CloudWatch agent Amazon sur AmazonEMR, la configuration par défaut publie les métriques système suivantes pour toutes les instances de votre cluster, sauf si vous [configurez l'agent différemment](#). Pour les définitions de chaque métrique, consultez la section [Mesures collectées par l' CloudWatch agent](#) dans le guide de CloudWatch l'utilisateur Amazon.

CPU

CPU métriques

- `cpu_usage_active`
- `cpu_usage_guest`
- `cpu_usage_guest_nice`
- `cpu_usage_idle`
- `cpu_usage_iowait`
- `cpu_usage_irq`

- `cpu_usage_nice`
- `cpu_usage_softirq`
- `cpu_usage_steal`
- `cpu_usage_system`
- `cpu_usage_user`

Disk

Métriques de disque

- `disk_free`
- `disk_total`
- `disk_used`
- `disk_used_percent`

Memory

Métrique de mémoire

- `mem_active`
- `mem_available`
- `mem_available_percent`
- `mem_free`
- `mem_inactive`
- `mem_total`
- `mem_used`
- `mem_used_percent`
- `mem_buffered`
- `mem_cached`

Network IO

Métriques d'E/S réseau

- `net_bytes_recv`

- `net_bytes_sent`
- `net_packets_recv`
- `net_packets_sent`

Process

Métriques de traitement

- `processes_running`
- `processes_total`

Swap

Métriques d'échange

- `swap_free`
- `swap_used`
- `swap_used_percent`

Par défaut, l'agent publie toutes les métriques CloudWatch dans l'espace de noms `CWAgent` personnalisé et sous le schéma `cluster.id,instance.id,node.type,service.name`. Suivez les étapes ci-dessous pour consulter les métriques par défaut publiées par l' CloudWatch agent pour Amazon EMR :

1. Accédez à la CloudWatch console Amazon.
2. Sélectionnez l'onglet Métriques, puis Toutes les métriques.
3. Sous Espaces de noms personnalisés, sélectionnez **CWAgent**. Sélectionnez ensuite le schéma : **cluster.id,instance.id,node.type,service.name**.
4. Continuez à interroger les métriques depuis l' CloudWatch interface selon votre cas d'utilisation.

Amazon EMR attache les étiquettes suivantes à chaque métrique : `["cluster.id", "instance.id", "node.type", "service.name"]`. Tenez compte des éléments suivants en ce qui concerne ces étiquettes :

- La valeur de l'`cluster.id` étiquette est l'ID du EMR cluster qui a produit la métrique. `j-123456789ABC` est un exemple de valeur pour l'étiquette `cluster.id`.

- La valeur de `instance.id` est l'ID de l'instance du EMR cluster qui a produit la métrique. `i-01bcf5f140f3355777` est un exemple de valeur pour `instance.id`.
- La valeur de `node.type` représente le type de nœud de l'instance du cluster qui a produit la métrique. Cette valeur peut être soit `master`, soit `task`.
- La valeur de l'étiquette `service.name` correspond au nom du service qui a produit la métrique. La valeur de `service.name` par défaut pour les métriques système par défaut est `system`.

Note

La EMR version 7.0.0 d'Amazon utilise un schéma d'étiquette métrique légèrement différent. `cluster.id` est `jobflow.id` et `node.type` n'est pas disponible. `instance.id` et `service.name` sont disponibles.

Configuration de CloudWatch l'agent pour Amazon EMR

Amazon EMR 7.0.0 et versions ultérieures incluent l'agent CloudWatch Amazon. Les sections suivantes décrivent comment configurer l'agent sur votre EMR cluster comme décrit dans [Créer un EMR cluster qui utilise l'agent CloudWatch Amazon](#).

Rubriques

- [Configurer CloudWatch l'agent pour Amazon EMR 7.1.0](#)
- [Configurer CloudWatch l'agent pour Amazon EMR 7.0.0](#)

Configurer CloudWatch l'agent pour Amazon EMR 7.1.0

À partir d'Amazon EMR 7.1.0, vous pouvez configurer l'agent CloudWatch Amazon pour utiliser des métriques système supplémentaires, ajouter des métriques d'application et modifier la destination des métriques en utilisant la EMR configuration API Amazon. Pour plus d'informations sur l'utilisation de la EMR configuration API pour configurer les applications de votre cluster, consultez la section [Configurer les applications](#).

Note

La version 7.1.0 ne prend en charge que le type de reconfiguration. `OVERWRITE` Pour plus d'informations sur les types de reconfiguration, consultez la section [Considérations relatives à la reconfiguration d'un groupe d'instances](#).

Rubriques

- [Schéma de configuration](#)
- [Exemples de configurations de métriques du système](#)
- [Exemples de configurations de métriques d'applications](#)
- [Exemple d'Amazon Managed Service pour Prometheus](#)

Schéma de configuration

`emr-metrics` possède les classifications suivantes :

- `emr-system-metrics`— configurez les métriques du système CPU, telles que le disque et la mémoire.
- `emr-hadoop-hdfs-datanode-metrics`— configurer les métriques Hadoop DataNode JMX
- `emr-hadoop-hdfs-namenode-metrics`— configurer les métriques Hadoop NameNode JMX
- `emr-hadoop-yarn-nodemanager-metrics`— configure les NodeManager JMX métriques Yarn
- `emr-hadoop-yarn-resourcemanager-metrics`— configure les ResourceManager JMX métriques Yarn
- `emr-hbase-master-metrics`— configurer les JMX métriques HBase principales
- `emr-hbase-region-server-metrics`— configurer les JMX métriques du serveur HBase régional
- `emr-hbase-rest-server-metrics`— configurer les JMX métriques HBase REST du serveur
- `emr-hbase-thrift-server-metrics`— configurer les métriques de HBase Thrift Server JMX

Les tableaux suivants décrivent les propriétés et les configurations disponibles pour toutes les classifications.

propriétés `emr-metrics`

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>metrics_destination</code>	Facultatif	Détermine si les métriques du cluster sont publiées sur Amazon CloudWatch ou sur Amazon Managed Service for Prometheus.	"CLOUDWATCH"	"CLOUDWATCH", "PROMETHEUS"	Cette propriété ne distingue pas les majuscules et minuscules. Par exemple, « Cloudwatch » est identique à « CLOUDWATCH ».
<code>prometheus_endpoint</code>	Facultatif	Si elle <code>metrics_destination</code> est définie sur PROMETHEUS « », cette propriété configure l'Amazon CloudWatch agent pour qu'il envoie des métriques au point de terminaison d'écriture à distance Amazon Managed	N/A	Tout service géré Amazon valide pour l'écriture à distance de Prometheus. URL Le format d'écriture à distance est <pre>https:// aps- worksp aces. <region> s.com/ wor kspaces/ <workspace_id> / api/v1/r</pre>	Ce champ est obligatoire s'il <code>metrics_destination</code> est défini sur « PROMETHEUS ». Le provisionnement échouera si vous ne fournissez pas de clé ou si la valeur est une chaîne vide.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
		Service for Prometheus fourni.		emote_wri te	

emr-system-metrics propriétés

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
metrics_collection_interval	Facultatif	Fréquence en secondes à laquelle les métriques sont collectées et publiées par l' CloudWatch agent.	« 60 »	Chaîne spécifiant le nombre de secondes. Accepte uniquement les nombres entiers.	Vous pouvez remplacer cette propriété par la metrics_collection_interval propriété provenant de groupes de mesures individuels.

emr-system-metrics configurations

cpu

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
metrics	Facultatif	La liste des CPU métriques	Voir les statistiques par défaut	Liste de noms de CPU	Si vous spécifiez une chaîne

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
		que l'agent doit collecter.	pour les CloudWatch agents auprès d'Amazon EMR	métriques valides séparés par des virgules, avec ou sans le <code>cpu_</code> préfixe, tels que <code>usage_active</code> <code>cpu_time_idle</code> . Consultez la section Mesures collectées par l'CloudWatch agent pour connaître les mesures valides.	vide, vous ne publiez aucune CPU métrique.
<code>metrics_collection_interval</code>	Facultatif	Fréquence en secondes à laquelle l'agent doit collecter et publier CPU les métriques .	La valeur du <code>globalmetrics_collection_interval</code> .	Chaîne spécifiant le nombre de secondes. Accepte uniquement les nombres entiers.	Cette valeur remplace la <code>metrics_collection_interval</code> propriété globale uniquement pour les CPU métriques.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>drop_original_metrics</code>	Facultatif	Liste des CPU mesures pour lesquelles il ne faut pas publier de mesures non agrégées.	Aucune CPU métrique non agrégée n'a été publiée.	Liste de mesures séparées par des CPU virgules qui sont également spécifiées dans la propriété <code>metrics</code> . Une chaîne vide signifie que toutes les CPU métriques doivent être publiées.	L' CloudWatch agent agrège toutes les métriques par ID de cluster, ID d'instance, type de nœud et nom de service. Par défaut, l' CloudWatch agent ne publie pas les métriques par ressource pour les métriques comportant plusieurs ressources.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>resources</code>	Facultatif	Détermine si l'agent publiera des métriques par cœur.	"*"	« * » active les métriques par cœur. « » désactive les métriques par cœur.	L' CloudWatch agent publie uniquement les métriques par cœur pour les CPU métriques qui ne sont pas enregistrées. <code>drop_original_metrics</code>

disk

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>metrics</code>	Facultatif	Liste des métriques de disque que l'agent doit collecter.	Voir les statistiques par défaut pour les CloudWatch agents auprès d'Amazon EMR	Liste séparée par des virgules de noms de métriques de disque valides avec ou sans le <code>disk_</code> préfixe, tel que <code>et.disk_total</code>	Si vous spécifiez une chaîne vide, vous ne publiez aucune métrique de disque.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
				used_percent Consultez la section Mesures collectées par l'agent CloudWatch pour connaître les mesures valides.	
metrics_collection_interval	Facultatif	Fréquence en secondes à laquelle l'agent doit collecter et publier les métriques du disque.	La valeur du globalmetrics_collection_interval.	Chaîne spécifiant le nombre de secondes. Accepte uniquement les nombres entiers.	Cette valeur remplace la metrics_collection_interval propriété globale uniquement pour les métriques de disque.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>drop_original_metrics</code>	Facultatif	Liste des mesures de disque pour lesquelles il ne faut pas publier de mesures non agrégées.	Aucune métrique de disque non agrégée n'a été publiée.	Liste séparée par des virgules de métriques de disque qui sont également spécifiées dans la propriété <code>metrics</code> . Une chaîne vide signifie que toutes les métriques du disque doivent être publiées.	L' CloudWatch agent agrège toutes les métriques par ID de cluster, ID d'instance, type de nœud et nom de service. Par défaut, l' CloudWatch agent ne publie pas les métriques par ressource pour les métriques comportant plusieurs ressources.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>resources</code>	Facultatif	Détermine si l'agent publiera per-mount-point des métriques.	"*"	« * » signifie tous les points de montage, « » signifie aucun point de montage ou une liste de points de montage séparés par des virgules. Par exemple, <code>"/,/emr"</code> .	L' CloudWatch agent publie uniquement per-mount-point les métriques de disque qui ne sont pas enregistrées dans <code>drop_original_metrics</code> .

diskio

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>metrics</code>	Facultatif	Liste des métriques d'E/S de disque que l'agent doit collecter.	Voir les statistiques par défaut pour les CloudWatch agents auprès d'Amazon EMR	Liste séparée par des virgules de noms de métriques d'E/S de disque valides avec ou sans le <code>diskio_</code> préfixe, tel que <code>et.diskio_re</code>	Si vous spécifiez une chaîne vide, vous ne publiez aucune métrique d'E/S sur disque.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
				ads writes Consultez la section Mesures collectées par l'agent CloudWatch pour connaître les mesures valides.	
<code>metrics_collection_interval</code>	Facultatif	Fréquence en secondes à laquelle l'agent doit collecter et publier les métriques d'E/S du disque.	La valeur de <code>globalmetrics_collection_interval</code> .	Chaîne spécifiant le nombre de secondes. Accepte uniquement les nombres entiers.	Cette valeur remplace la propriété <code>metrics_collection_interval</code> globalement pour les métriques d'E/S de disque.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>drop_original_metrics</code>	Facultatif	Liste des mesures d'E/S sur disque pour lesquelles il ne faut pas publier de mesures non agrégées.	Aucune métrique d'E/S de disque non agrégée n'a été publiée.	Liste séparée par des virgules de métriques d'E/S sur disque qui sont également spécifiées dans la propriété <code>metrics</code> . Une chaîne vide signifie que toutes les métriques d'E/S du disque doivent être publiées.	L' CloudWatch agent agrège toutes les métriques par ID de cluster, ID d'instance, type de nœud et nom de service. Par défaut, l' CloudWatch agent ne publie pas les métriques par ressource pour les métriques comportant plusieurs ressources.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>resources</code>	Facultatif	Détermine si l'agent publiera des métriques par appareil.	"*"	« * » désigne tous les périphériques de stockage, « » signifie aucun périphérique de stockage ou une liste de noms de périphériques séparés par des virgules. Par exemple, "nvme0n1, nvme1n1" .	L' CloudWatch agent publie uniquement les métriques par appareil pour les métriques d'E/S sur disque qui ne sont pas introduites. <code>drop_original_metrics</code>

mem

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>metrics</code>	Facultatif	Liste des métriques de mémoire que l'agent doit collecter.	Voir les statistiques par défaut pour les CloudWatch agents auprès d'Amazon EMR	Liste séparée par des virgules de noms de métriques de mémoire valides avec ou sans le <code>mem_</code> préfixe,	Spécifier une chaîne vide signifie ne publier aucune métrique de mémoire.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
				tels que <code>et.mem_available_percent</code> . Consultez la section Mesures collectées par l'agent CloudWatch pour connaître les mesures valides.	
<code>metrics_collection_interval</code>	Facultatif	Fréquence en secondes à laquelle l'agent doit collecter et publier les métriques de mémoire.	La valeur du <code>globalmetrics_collection_interval</code> .	Chaîne spécifiant le nombre de secondes. Accepte uniquement les nombres entiers.	Cette valeur remplace la <code>metrics_collection_interval</code> propriété globale uniquement pour les métriques de mémoire.

net

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>metrics</code>	Facultatif	Liste des métriques réseau que l'agent doit collecter.	Voir les statistiques par défaut pour les CloudWatch agents auprès d'Amazon EMR	Liste séparée par des virgules de noms de métriques réseau valides avec ou sans le <code>net_</code> préfixe, tels que <code>et.net_packets_received</code> . Consultez la section Mesures collectées par l'agent CloudWatch pour connaître les mesures valides.	Si vous spécifiez une chaîne vide, vous ne publiez aucune métrique réseau.
<code>metrics_collection_interval</code>	Facultatif	Fréquence en secondes à laquelle l'agent doit collecter et publier les	La valeur du <code>globalmetrics_collection_interval</code> .	Chaîne spécifiant le nombre de secondes. Accepte uniquement	Cette valeur remplace la <code>metrics_collection_interval</code> propriété globale

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
		métriques du réseau.		les nombres entiers.	uniquement pour les métriques du réseau.
drop_original_metrics	Facultatif	Liste des mesures du réseau pour lesquelles il ne faut pas publier de mesures non agrégées.	Aucune métrique de réseau non agrégée n'a été publiée.	Liste séparée par des virgules de métriques réseau qui sont également spécifiées dans la propriété metrics. Une chaîne vide signifie que toutes les métriques du réseau sont publiées.	L' CloudWatch agent agrège toutes les métriques par ID de cluster, ID d'instance, type de nœud et nom de service. Par défaut, l' CloudWatch agent ne publie pas les métriques par ressource pour les métriques comportant plusieurs ressources.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>resources</code>	Facultatif	Détermine si l'agent publiera des métriques par interface.	"*"	« * » signifie toutes les interfaces réseau, « » signifie aucune interface réseau ou une liste de noms d'interfaces séparés par des virgules. Par exemple, "eth0,eth1" .	L' CloudWatch agent publie uniquement les métriques par interface pour les métriques réseau qui ne sont pas intégrées. <code>drop_original_metrics</code>

netstat

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>metrics</code>	Facultatif	Liste des statistiques réseau que l'agent doit collecter.	Voir les statistiques par défaut pour les CloudWatch agents auprès d'Amazon EMR	Liste séparée par des virgules de noms de métriques de mémoire valides avec ou sans le <code>netstat_</code> préfixe, tels que et.	Spécifier une chaîne vide signifie ne publier aucune métrique statistique du réseau.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
				<p>tcp_liste n netstat_u dp_socket</p> <p>Consultez la section Mesures collectées par l'CloudWatch agent pour connaître les mesures valides.</p>	
metrics_collection_interval	Facultatif	Fréquence en secondes à laquelle l'agent doit collecter et publier les statistiques du réseau.	La valeur du globalmetrics_collection_interval.	Chaîne spécifiant le nombre de secondes. Accepte uniquement les nombres entiers.	Cette valeur remplace la metrics_collection_interval propriété globale uniquement pour les métriques statistiques du réseau.

processes

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>metrics</code>	Facultatif	Liste des métriques de processus que l'agent doit collecter.	Voir les statistiques par défaut pour les CloudWatch agents auprès d'Amazon EMR	Liste séparée par des virgules de noms de métriques de mémoire valides avec ou sans le processus _ préfixe, tels que <code>et.processes_running</code> <code>total</code> . Consultez la section Mesures collectées par l'CloudWatch agent pour connaître les mesures valides.	Spécifier une chaîne vide signifie ne publier aucune métrique de processus.
<code>metrics_collection_interval</code>	Facultatif	Fréquence en secondes à laquelle l'agent doit collecter et publier les métriques	La valeur du <code>globalmetrics_collection_interval</code> .	Chaîne spécifiant le nombre de secondes. Accepte uniquement	Cette valeur remplace la <code>metrics_collection_interval</code> propriété globale

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
		des processus du système.		les nombres entiers.	uniquement pour les métriques de processus du système.

swap

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
<code>metrics</code>	Facultatif	La liste des métriques de swap que l'agent doit collecter.	Voir les statistiques par défaut pour les CloudWatch agents auprès d'Amazon EMR	Liste séparée par des virgules de noms de métriques de mémoire valides avec ou sans le <code>swap_</code> préfixe, tels que <code>swap_free</code> <code>used_percent</code> Consultez la section Mesures collectées par l'CloudWatch agent pour	Spécifier une chaîne vide signifie ne publier aucune métrique de swap.

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	Remarques
				connaître les mesures valides.	
<code>metrics_collection_interval</code>	Facultatif	Fréquence en secondes à laquelle l'agent doit collecter et publier les métriques de swap.	La valeur du <code>globalmetrics_collection_interval</code> .	Chaîne spécifiant le nombre de secondes. Accepte uniquement les nombres entiers.	Cette valeur remplace la <code>metrics_collection_interval</code> propriété globale uniquement pour les métriques de swap.

emr-hadoop-hdfs-datanode-propriétés des métriques

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	
<code><custom_bean_name></code>	Facultatif	N/A	MBeanSet CloudWatch agent doit collecter des métriques, telles que <code>Hadoop:service=DataNode,name=DataNodeActivity</code> . Vous	Chaîne contenant la liste séparée par des virgules des métriques associées au MBean. Par exemple, <code>BlocksCached,BlockRead</code> .	

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles
			trouvez des exemples de MBean noms et les statistiques correspondantes dans les JMX YAML fichiers d'exemple pour Amazon EMR version 7.0.	
<code>otel.metric.export.interval</code>	Facultatif	Fréquence, en millisecondes, de collecte des métriques Hadoop. DataNode	« 60 000 »	Chaîne spécifiant le nombre de millisecondes. Accepte uniquement les nombres entiers.

emr-hadoop-hdfs-namenode-propriétés des métriques

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles
<code><custom_bean_name></code>	Facultatif	N/A	MBeanCet CloudWatch agent doit collecter des	Chaîne contenant la liste séparée par des

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles
			<p>métriques , telles queHadoop:service=NameNode, name=FSNamesystem . Vous trouverez des exemples de MBean noms et les statistiques correspondantes dans les JMXAMLFichiers d'exemple pour Amazon EMR version 7.0.</p>	<p>virgules des métriques associées au. MBean Par exemple, BlockCapacity, CapacityUsedGB .</p>
otel.metric.export.interval	Facultatif	Fréquence, en millisecondes, de collecte des métriques Hadoop. NameNode	« 60 000 »	Chaîne spécifiant le nombre de millisecondes. Accepte uniquement les nombres entiers.

emr-hadoop-yarn-nodemanager-propriétés des métriques

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles
<code><custom_bean_name></code>	Facultatif	N/A	MBeanCet CloudWatch agent doit collecter des métriques, telles queHadoop:service=NodeManager, name=NodeManagerMetrics . Vous trouverez des exemples de MBean noms et les statistiques correspondantes dans les JMXAMLFichiers d'exemple pour Amazon EMR version 7.0.	Chaîne contenant la liste séparée par des virgules des métriques associées au MBean. Par exemple, MaxCapacity, AllocatedGB .
<code>otel.metric.export.interval</code>	Facultatif	Fréquence, en millisecondes, de collecte des	« 60 000 »	Chaîne spécifiant le nombre de milliseco

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	
		métriques Hadoop. YARN NodeManager		ndes. Accepte uniquement les nombres entiers.	

emr-hadoop-yarn-resourcemanager-propriétés des métriques

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	
<code><custom_bean_name></code>	Facultatif	N/A	MBeanCet CloudWatch agent doit collecter des métriques, telles queHadoop:service=ResourceManager,name=PartitionQueueMetrics. Vous trouverez des exemples de MBean noms et les statistiques correspondantes dans les JMXYAMLfi	Chaîne contenant la liste séparée par des virgules des métriques associées au. MBean Par exemple, MaxCapacity,MaxCapacityVCores.	

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles
			chiers d'exemple pour Amazon EMR version 7.0.	
<code>otel.metric.export.interval</code>	Facultatif	Fréquence, en millisecondes, de collecte des métriques Hadoop. YARN ResourceManager	« 60 000 »	Chaîne spécifiant le nombre de millisecondes. Accepte uniquement les nombres entiers.

emr-hbase-master-metrics propriétés

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles
<code><custom_bean_name></code>	Facultatif	N/A	MBeanCet CloudWatch agent doit collecter des métriques, telles queHadoop:service=HBase, name=Master, sub=AssignmentManager	Chaîne contenant la liste séparée par des virgules des métriques associées au. MBean Par exemple, AssignFailedCount, AssignSub

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	
			. Vous trouverez des exemples de MBean noms et les statistiques correspondantes dans les JMX YAML fichiers d'exemple pour Amazon EMR version 7.0.	mittedCount .	
otel.metric.export.interval	Facultatif	Fréquence, en millisecondes, de collecte HBase des métriques principales.	« 60 000 »	Chaîne spécifiant le nombre de millisecondes. Accepte uniquement les nombres entiers.	

emr-hbase-region-server-propriétés des métriques

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	
<i><custom_mbean_name></i>	Facultatif	N/A	MBeanSet CloudWatch agent doit	Chaîne contenant la liste séparée	

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles
			collecter des métriques , telles queHadoop:service=HBase,name=RegionServer,sub=IPC . Vous trouverez des exemples de MBean noms et les statistiques correspondantes dans les JMXYAMLfichiers d'exemple pour Amazon EMR version 7.0.	par des virgules des métriques associées au. MBean Par exemple, numActiveHandler,numActivePriorityHandler .
otel.metric.export.interval	Facultatif	Fréquence, en millisecondes, de collecte des métriques du serveur HBase régional.	« 60 000 »	Chaîne spécifiant le nombre de millisecondes. Accepte uniquement les nombres entiers.

emr-hbase-rest-server-propriétés des métriques

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles
<code><custom_bean_name></code>	Facultatif	N/A	MBeanCet CloudWatch agent doit collecter des métriques, telles queHadoop:service=HBase, name=REST. Vous trouverez des exemples de MBean noms et les statistiques correspondantes dans les JMXAML fichiers d'exemple pour Amazon EMR version 7.0.	Chaîne contenant la liste séparée par des virgules des métriques associées au MBean. Par exemple, <code>successfulPut, successfulScanCount</code> .
<code>otel.metric.export.interval</code>	Facultatif	Fréquence, en millisecondes, de collecte des métriques du serveur HBase Rest.	« 60 000 »	Chaîne spécifiant le nombre de millisecondes. Accepte uniquement

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	
				les nombres entiers.	

emr-hbase-thrift-server-propriétés des métriques

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles	
<code><custom_bean_name></code>	Facultatif	N/A	MBeanCet CloudWatch agent doit collecter des métriques, telles queHadoop:service=HBase,name=Thrift,sub=ThriftOne. Vous trouverez des exemples de MBean noms et les statistiques correspondantes dans les JMXAMLfi chiers d'exemple pour Amazon	Chaîne contenant la liste séparée par des virgules des métriques associées au. MBean Par exemple, BatchGet_max, BatchGet_mean.	

Propriété	Obligatoire	Description	Valeur par défaut	Valeurs possibles
			EMR version 7.0.	
<code>otel.metric.export.interval</code>	Facultatif	Fréquence, en millisecondes, de collecte des métriques du serveur HBase Thrift.	« 60 000 »	Chaîne spécifiant le nombre de millisecondes. Accepte uniquement les nombres entiers.

Exemples de configurations de métriques du système

L'exemple suivant montre comment configurer l' CloudWatch agent pour arrêter d'exporter toutes les métriques du système.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
]
```

L'exemple suivant configure l' CloudWatch agent pour exporter les métriques système par défaut. Cela constitue un moyen rapide de réinitialiser l'agent pour qu'il n'exporte les métriques système par défaut que si vous les avez déjà reconfigurées au moins une fois. Cette réinitialisation supprime également toutes les métriques d'application qui ont été reconfigurées auparavant.


```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": []
  }
]
```

L'exemple suivant configure le cluster pour exporter les disk métriques cpumem,, et.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {
          "metrics_collection_interval": "20"
        },
        "Configurations": [
          {
            "Classification": "cpu",
            "Properties": {
              "metrics": "cpu_usage_guest,cpu_usage_idle",
              "metrics_collection_interval": "30",
              "drop_original_metrics": "cpu_usage_guest"
            }
          },
          {
            "Classification": "mem",
            "Properties": {
              "metrics": "mem_active"
            }
          },
          {
            "Classification": "disk",
            "Properties": {
              "metrics": "disk_used_percent",
              "resources": "/,/mnt",
              "drop_original_metrics": ""
            }
          }
        ]
      }
    ]
  }
]
```

```

    ]
  }
]
}
]

```

L'exemple de configuration précédent possède les propriétés suivantes :

- Toutes les 30 secondes, l'agent collecte la `cpu_guest` métrique pour tous CPUs. Vous pouvez trouver la métrique agrégée sous l'espace de CloudWatch noms `CWAgent > cluster.id, instance.id, node.type, service.name`.
- Toutes les 30 secondes, l'agent collecte la `cpu_idle` métrique pour tous CPUs. Vous pouvez trouver la métrique agrégée sous l'espace de CloudWatch noms `CWAgent > cluster.id, instance.id, node.type, service.name`. L'agent collecte également les métriques par processeur. Vous pouvez les trouver dans le même espace de noms. L'agent collecte cette métrique parce que la `drop_original_metrics` propriété ne contient pas `cpu_idle`, il n'ignore donc pas la métrique.
- Toutes les 20 secondes, l'agent collecte la `mem_active` métrique. Vous pouvez trouver la métrique agrégée sous l'espace de CloudWatch noms `CWAgent > cluster.id, instance.id, node.type, service.name`.
- Toutes les 20 secondes, l'agent collecte les `disk_used_percent` métriques pour les montages `/mnt` sur disque `/` et. Vous pouvez trouver les métriques agrégées sous l'espace de CloudWatch noms `CWAgent > cluster.id, instance.id, node.type, service.name`. L'agent collecte également les métriques par montage. Vous pouvez les trouver dans le même espace de noms. L'agent collecte cette métrique parce que la `drop_original_metrics` propriété ne contient pas `disk_used_percent`, il n'ignore donc pas la métrique.

Exemples de configurations de métriques d'applications

L'exemple suivant configure l' CloudWatch agent pour arrêter d'exporter les métriques pour le service Hadoop Namenode.

```

[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {

```

```

    "Classification": "emr-hadoop-hdfs-namenode-metrics",
    "Properties": {},
    "Configurations": []
  }
]

```

L'exemple suivant configure un cluster pour exporter les métriques d'applications Hadoop.

```

[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-hadoop-hdfs-namenode-metrics",
        "Properties": {
          "Hadoop:service=NameNode,name=FSNamesystem": "BlockCapacity,CapacityUsedGB",
          "otel.metric.export.interval": "20000"
        },
        "Configurations": []
      },
      {
        "Classification": "emr-hadoop-hdfs-datanode-metrics",
        "Properties": {
          "Hadoop:service=DataNode,name=JvmMetrics": "MemNonHeapUsedM",
          "otel.metric.export.interval": "30000"
        },
        "Configurations": []
      },
      {
        "Classification": "emr-hadoop-yarn-resourcemanager-metrics",
        "Properties": {
          "Hadoop:service=ResourceManager,name=CapacitySchedulerMetrics":
"AllocateNumOps,NodeUpdateNumOps"
        },
        "Configurations": []
      }
    ]
  }
]

```

L'exemple précédent possède les propriétés suivantes :

- Toutes les 20 secondes, l'agent collecte les CapacityUsedGB métriques BlockCapacity et auprès des instances exécutant le service Hadoop Namenode.
- Toutes les 30 secondes, l'agent collecte les MemNonHeapUsedM métriques des instances exécutant le service Hadoop Datanode.
- Toutes les 30 secondes, l'agent collecte les NodeUpdateNumOps métriques AllocateNumOps et à partir des instances qui exécutent Hadoop YARN ResourceManaager.

Exemple d'Amazon Managed Service pour Prometheus

L'exemple suivant montre comment configurer l' CloudWatch agent pour exporter des métriques vers Amazon Managed Service for Prometheus.

Si vous exportez actuellement des métriques vers Amazon Managed Service for Prometheus et que vous souhaitez reconfigurer les métriques pour le cluster et continuer à exporter des métriques vers Amazon Managed Service for Prometheus, vous devez inclure les propriétés et. `metrics_destination` `prometheus_endpoint`

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "prometheus",
      "prometheus_endpoint": "http://amp-workspace/api/v1/remote_write"
    },
    "Configurations": []
  }
]
```

Pour utiliser l' CloudWatch agent vers lequel exporter des métriques CloudWatch, utilisez l'exemple suivant.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "cloudwatch"
    },
  },
]
```

```
"Configurations": []  
}  
]
```

Note

L' CloudWatch agent dispose d'un exportateur Prometheus qui renomme certains attributs. Pour les étiquettes de métriques par défaut, Amazon Managed Service for Prometheus utilise des caractères de soulignement à la place des points utilisés par Amazon. CloudWatch Si vous utilisez Amazon Managed Grafana pour visualiser les métriques par défaut dans Amazon Managed Service for Prometheus, les libellés apparaissent sous la forme,, et. `cluster_id instance_id node_type service_name`

Configurer CloudWatch l'agent pour Amazon EMR 7.0.0

Vous pouvez configurer l' CloudWatch agent Amazon pour utiliser des métriques système supplémentaires au-delà de celles fournies [par la configuration de l' CloudWatch agent par défaut](#). La configuration de la version 7.0.0 requiert l'utilisation d'actions d'amorçage, dont nous avons fourni des exemples dans les sections suivantes. Dans une prochaine version, Amazon EMR fournira des options de configuration supplémentaires via Amazon EMRAPI.

Rubriques

- [Configurer des métriques système supplémentaires avec Amazon EMR 7.0.0](#)
- [Configurer les métriques des applications avec Amazon EMR 7.0.0](#)
- [Configurer Amazon Managed Service pour Prometheus en tant que stockage cloud pour les métriques avec Amazon 7.0.0 EMR](#)

Configurer des métriques système supplémentaires avec Amazon EMR 7.0.0

Suivez les étapes suivantes pour configurer l'agent afin qu'il utilise un ensemble différent de métriques système dans Amazon EMR 7.0.0 :

1. Créez ou choisissez un compartiment dans votre compte Amazon S3 dans lequel vous souhaitez stocker les fichiers de configuration qui spécifient les métriques de l' CloudWatch agent.
2. Créez le fichier de configuration `emr-amazon-cloudwatch-agent.json` en spécifiant vos métriques préférées. Pour ce faire, utilisez l'une des méthodes expliquées dans la section

[Création du fichier de configuration de l' CloudWatch agent](#). Pour plus d'informations sur la structure du fichier de configuration de l' CloudWatch agent, consultez la section [Création ou modification manuelle du fichier de configuration de l' CloudWatch agent](#) dans le guide de CloudWatch l'utilisateur Amazon.

3. Accédez ensuite au aws-emr-utilities dépôt GitHub et téléchargez les scripts de métriques système suivants :
 - [install_system_metrics_launcher.sh](#) – Script qui télécharge et exécute `install_system_metrics.sh` en arrière-plan pour que le nœud puisse terminer l'amorçage.
 - [install_system_metrics.sh](#)— Script qui attend la fin du démarrage de l'instance sur laquelle il s'exécute, puis télécharge et applique la configuration contenue dans le fichier. JSON
4. Ouvrez chaque fichier SH et remplacez *my-s3-bucket* par le nom de votre compartiment indiqué à l'étape 1.
5. Téléchargez les un JSON et deux fichiers SH dans votre compartiment S3.
6. Vous pouvez désormais accéder à la EMR console Amazon et créer un nouveau cluster avec l' CloudWatch agent. Sous EMRActivé EC2 dans le volet de navigation de gauche, sélectionnez Clusters, puis Créer un cluster.
7. Dans la section Nom et applications, choisissez une EMR version Amazon 7.0.0 ou supérieure.
8. Sous Bundle d'applications, sélectionnez le bundle ou les applications que vous souhaitez installer sur votre cluster, et incluez Amazon CloudWatch Agent dans vos sélections.
9. Dans la section Actions d'amorçage, sélectionnez Ajouter.
 - Pour le Nom, insérez **install_system_metrics_launcher.sh**.
 - Pour l'Emplacement du script, insérez **s3://my-s3-bucket/install_system_metrics_launcher.sh**. Remplacez *my-s3-bucket* par le chemin de votre compartiment S3.
 - Laissez le bloc Arguments vide.
10. Sélectionnez Ajouter une action d'amorçage.
11. Continuez à créer le cluster pour répondre à vos besoins en matière de charge de travail.

Lorsque votre cluster est lancé, l' CloudWatch agent publie les métriques système que vous avez spécifiées dans le fichier de configuration sur CloudWatch.

Configurer les métriques des applications avec Amazon EMR 7.0.0

Vous pouvez configurer l' CloudWatch agent Amazon pour publier des métriques d'application pour HDFS et YARN en plus des métriques du système. Pour configurer l'agent de sorte à publier des métriques d'application, procédez comme suit :

1. Créez ou choisissez un compartiment dans votre compte Amazon S3 dans lequel vous souhaitez stocker les fichiers de configuration qui spécifient les métriques de l' CloudWatch agent.
2. Accédez ensuite au [aws-emr-utilities](#) dépôt GitHub et téléchargez les scripts suivants :
 - [install_app_metrics_launcher.sh](#) – Script qui télécharge et exécute `install_app_metrics.sh` en arrière-plan pour que le nœud puisse terminer l'amorçage.
 - [install_app_metrics.sh](#)— Script qui attend la fin du démarrage de l'instance sur laquelle il s'exécute, puis télécharge et applique la configuration dans les YAML fichiers que vous téléchargerez lors d'une prochaine étape.
3. Ouvrez chaque fichier et remplacez *my-s3-bucket* par le nom de votre compartiment indiqué à l'étape 1.
4. Téléchargez ensuite les fichiers de YAML mappage suivants. Pour plus d'informations sur la structure de ces YAML fichiers, consultez [javaagent](#) le OpenTelemetry Instrumentation for Java GitHub dépôt.
 - [datanode-metrics.yaml](#) –Configuration des métriques Hadoop DataNode.
 - [namenode-metrics.yaml](#) –Configuration des métriques Hadoop NameNode.
 - [nodemanager-metrics.yaml](#) –Configuration des métriques Yarn NodeManager.
 - [resourcemanager-metrics.yaml](#) –Configuration des métriques Yarn ResourceManager.
5. Téléchargez les deux fichiers SH et YAML les quatre fichiers dans votre compartiment S3.
6. Vous pouvez désormais accéder à la EMR console Amazon et créer un nouveau cluster avec l' CloudWatch agent. Sous EMRActivé EC2 dans le volet de navigation de gauche, sélectionnez Clusters, puis Créer un cluster.
7. Dans la section Nom et applications, choisissez une EMR version Amazon 7.0.0 ou supérieure.
8. Sous Ensemble d'applications, sélectionnez le bundle ou le groupe personnalisé d'applications que vous souhaitez installer sur votre cluster, et incluez CloudWatch l'agent dans vos sélections.
9. Dans la section Actions d'amorçage, sélectionnez Ajouter.
 - Pour le Nom, insérez **install_app_metrics_launcher.sh**.

- Pour l'Emplacement du script, insérez **s3://my-s3-bucket/install_app_metrics_launcher.sh**. Remplacez *my-s3-bucket* par le chemin de votre compartiment S3.
 - Laissez le bloc Arguments vide.
10. Sélectionnez Ajouter une action d'amorçage.
 11. Continuez à créer le cluster pour répondre à vos besoins en matière de charge de travail.

Lorsque votre cluster est lancé, l' CloudWatch agent publie les métriques d'application que vous avez spécifiées ainsi que les métriques système dans CloudWatch.

Configurer Amazon Managed Service pour Prometheus en tant que stockage cloud pour les métriques avec Amazon 7.0.0 EMR

Vous pouvez configurer l' CloudWatch agent Amazon pour qu'il publie les métriques sur Amazon Managed Service for Prometheus au lieu de. CloudWatch

Note

Vous pouvez publier les métriques de l' CloudWatch agent Amazon sur Amazon Managed Service for Prometheus ou sur CloudWatch Amazon, mais vous ne pouvez pas publier les métriques sur les deux services pour le même cluster.

Pour configurer l'agent afin qu'il publie des métriques sur Amazon Managed Service for Prometheus, vous devez ajouter `aps:RemoteWrite` AWS Identity and Access Management l'autorisation IAM () au profil d'instance Amazon pour EC2 Amazon. EMR L'exemple de politique suivant contient les autorisations requises :

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "aps:RemoteWrite",
      "Resource": "*"
    }
  ]
}
```


Utiliser l' CloudWatch agent sur un EMR cluster pour publier des métriques sur Amazon Managed Service for Prometheus

Une fois que la politique de service dispose des autorisations appropriées, suivez les étapes suivantes pour lancer un cluster qui utilise l' CloudWatch agent pour publier des métriques sur Amazon Managed Service for Prometheus.

1. Utilisez le AWS Management Console ou AWS CLI pour créer un espace de travail Amazon Managed Service pour Prometheus. Pour plus d'informations, voir la rubrique [Création d'un Workspace](#) du Guide de l'utilisateur Amazon Managed Service for Prometheus.
2. Créez ou choisissez un compartiment dans votre compte Amazon S3 dans lequel vous souhaitez stocker les fichiers de lancement qui spécifient Amazon Managed Service for Prometheus en tant qu'espace de stockage cloud.
3. Accédez ensuite au `aws-emr-utilities` dépôt GitHub et téléchargez les scripts suivants :
 - [add_prometheus_endpoint_launcher.sh](#) – Script qui télécharge et exécute `add_prometheus_endpoint.sh` en arrière-plan pour que le nœud puisse terminer l'amorçage.
 - [add_prometheus_endpoint.sh](#)— Un script qui attend la fin du démarrage de l'instance sur laquelle il s'exécute, puis configure l' CloudWatch agent à publier sur le point de terminaison Amazon Managed Service for Prometheus que vous fournissez comme argument lorsque vous lancez votre cluster.
4. Ouvrez chaque fichier et remplacez `my-s3-bucket` par le nom de votre compartiment indiqué à l'étape 2.
5. Utilisez le AWS CLI pour créer un EMR cluster avec l'action `add_prometheus_endpoint_launcher.sh` bootstrap. Dans la commande suivante, remplacez `my-s3-bucket` par le compartiment contenant l'action d'amorçage et remplacez `managedpro-remote-write-workspace-url` par le point de terminaison d'écriture à distance pour votre Workspace Amazon Managed Service for Prometheus. Assurez-vous de spécifier une étiquette de EMR sortie Amazon égale `emr-7.0.0` ou supérieure.

```
aws emr create-cluster --name managedpro-cluster \  
  --release-label emr-7.0.0 \  
  --applications Name=Hadoop Name=AmazonCloudWatchAgent \  
  --ec2-attributes KeyName=myKey --instance-type m7g.2xlarge \  
  --instance-count 3 --use-default-roles
```

```
--bootstrap-actions Name='Add Prometheus Endpoint',Path=s3://my-s3-bucket/  
add_prometheus_endpoint_launcher.sh,Args='managedpro-remote-write-workspace-url'
```

Lorsque votre cluster est lancé, l' CloudWatch agent publie les métriques qu'il collecte sur Amazon Managed Service for Prometheus.

Utiliser Amazon Managed Service for Prometheus en tant que source de données pour Amazon Managed Grafana

Une fois qu'Amazon EMR a publié les métriques du cluster sur Amazon Managed Service for Prometheus, vous pouvez suivre les étapes suivantes pour visualiser les métriques avec Amazon Managed Grafana :

1. Utilisez le AWS Management Console pour créer un espace de travail et un utilisateur Amazon Managed Grafana disposant des autorisations appropriées. Pour plus d'informations, voir la rubrique [Création d'un Workspace](#) du Guide de l'utilisateur Amazon Managed Grafana.
2. Ajoutez votre Workspace Amazon Managed Service for Prometheus en tant que source de données pour Amazon Managed Grafana. Pour plus d'informations, voir la rubrique [Utiliser la configuration de la source de données AWS pour ajouter Amazon Managed Service for Prometheus en tant que source de données](#) du Guide de l'utilisateur Amazon Managed Grafana.

Note

L' CloudWatch agent dispose d'un exportateur Prometheus qui renomme certains attributs. Pour les étiquettes de métriques par défaut, Amazon Managed Service for Prometheus utilise des caractères de soulignement à la place des points utilisés par Amazon. CloudWatch Si vous utilisez Amazon Managed Grafana pour visualiser les métriques par défaut dans Amazon Managed Service for Prometheus, les étiquettes s'affichent sous la forme `jobflow_id`, `instance_id` et `service_name`. De plus, toutes les métriques d'application publiées par l' CloudWatch agent sur Amazon Managed Service for Prometheus utilisent l'étiquette à la place de `job.service_name`. Cependant, les métriques système continuent d'utiliser l'étiquette `service_name`.

Considérations et restrictions

- L' CloudWatch agent Amazon natif est disponible pour les clusters que vous créez avec les EMR versions 7.0.0 et supérieures d'Amazon.
- La configuration de CloudWatch l'agent dans Amazon EMR 7.0.0 nécessite l'utilisation d'actions [bootstrap](#). Dans une prochaine version, Amazon EMR fournira des options de configuration supplémentaires via Amazon EMR API.
- Vous ne pouvez pas installer l'EMR CloudWatch agent Amazon si vous l' CloudWatch avez déjà déployé par une autre méthode, telle que AWS Systems Manager l'agent (SSMagent) dans la région où vous créez votre cluster. Cette opération entraînera l'arrêt de votre cluster avec des erreurs.

CloudWatch historique des versions de l'agent

Le tableau suivant répertorie la version AmazonCloudWatchAgent incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

AmazonCloudWatchAgent informations sur la version

Étiquette Amazon EMR Release	AmazonCloudWatchAgent Version	Composants installés avec AmazonCloudWatchAgent
emr-7.2.0	1,300032,2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resour

Étiquette Amazon EMR Release	AmazonCloudWatchAgent Version	Composants installés avec AmazonCloudWatchAgent
		cemanager, hadoop-yarn-timeline-server
emr-7.1.0	1,300032,2	adot-java-agent, emrfs, emr-amazon-cloudwatch-agent, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-7.0.0	1,300031,1	adot-java-agent, emr-amazon-cloudwatch-agent

Delta Lake

Delta Lake est un framework de couche de stockage pour les architectures Lakehouse généralement construites sur Amazon S3. Avec les EMR versions 6.9.0 et supérieures d'Amazon, vous pouvez utiliser [Apache Spark 3.x](#) sur des EMR clusters Amazon dotés de tables Delta Lake. Pour plus d'informations sur les cabanes au bord du lac Delta, consultez le site <https://delta.io/>.

Le tableau suivant répertorie la version de Delta incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Delta.

Pour la version des composants installés avec Delta dans cette version, voir la [version 7.2.0 Versions des composants](#).

Informations sur la version Delta pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Delta	Composants installés avec Delta
emr-7.2.0	Delta 3.1.0	Not available.

Le tableau suivant répertorie la version de Delta incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Delta.

Pour connaître la version des composants installés avec Delta dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version Delta pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Delta	Composants installés avec Delta
emr-6.15.0	Delta 2.4.0	Not available.

Présentation de Delta Lake

Delta Lake est un projet open source qui permet de mettre en œuvre des architectures de lacs de données modernes généralement basées sur Amazon S3. Delta Lake propose les fonctions suivantes :

- Transactions atomiques, cohérentes, isolées, durables (ACID) sur Spark. Les lecteurs bénéficient d'une vue cohérente du tableau lors d'une tâche Spark.
- Gestion évolutive des métadonnées avec traitement distribué par Spark.
- Combine les cas de streaming et d'utilisation par lots avec la même table Delta.
- Application automatique du schéma pour éviter les mauvais enregistrements lors de l'ingestion de données.
- Voyage dans le temps avec gestion des versions des données.
- Prend en charge les opérations de fusion, de mise à jour et de suppression pour les cas d'utilisation complexes tels que la capture des données de modification (CDC), le streaming upserts, etc.

Utilisation d'un cluster sur lequel Delta Lake est installé

Rubriques

- [Utiliser un cluster Delta Lake avec Flink](#)
- [Utilisez un cluster Delta Lake avec Trino](#)
- [Utiliser un cluster Delta Lake avec Spark](#)
- [Utilisez un cluster Delta Lake avec Spark et AWS Glue](#)

Utiliser un cluster Delta Lake avec Flink

À partir EMR de la version 6.11 d'Amazon, vous pouvez utiliser Delta Lake avec votre cluster Flink. Les exemples suivants utilisent le AWS CLI pour travailler avec Delta Lake sur un cluster Amazon EMR Flink.

Note

Amazon EMR prend en charge le Flink DataStream API lorsque vous utilisez Delta Lake avec un cluster Flink.

Création d'un cluster Delta Lake

1. Créez un fichier, `delta_configurations.json`, contenant les éléments suivants :

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}}]
```

2. Créez un cluster avec la configuration suivante. Remplacez exemple Amazon S3 bucket path et subnet ID par les vôtres.

```
aws emr create-cluster
--release-label emr-6.11.0
--applications Name=Flink
--configurations file://delta_configurations.json
--region us-east-1 --name My_Spark_Delta_Cluster
--log-uri s3://DOC-EXAMPLE-BUCKET/
--instance-type m5.xlarge
--instance-count 3
--service-role EMR_DefaultRole_V2
--ec2-attributes
InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Initialisation d'une session Flink yarn

Pour initialiser une session Flink yarn, exécutez la commande suivante :

```
flink-yarn-session -d
```

Créer une tâche Flink avec Delta Lake

Les exemples suivants montrent comment utiliser sbt ou Maven pour créer votre tâche Flink avec Delta Lake.

sbt

[sbt](#) est un outil de construction pour Scala que vous pouvez utiliser avec peu ou pas de configuration lorsque vous avez de petits projets.

```
libraryDependencies += Seq(
  "io.delta" %% "delta-flink" % deltaConnectorsVersion % "provided",
  "io.delta" %% "delta-standalone" % deltaConnectorsVersion % "provided",
  "org.apache.flink" %% "flink-clients" % flinkVersion % "provided",
  "org.apache.flink" %% "flink-parquet" % flinkVersion % "provided",
  "org.apache.hadoop" % "hadoop-client" % hadoopVersion % "provided",
```

```
"org.apache.flink" % "flink-table-common" % flinkVersion % "provided",
"org.apache.flink" %% "flink-table-runtime" % flinkVersion % "provided")
```

Maven

[Maven](#) est un outil d'automatisation de build open source de l'Apache Software Foundation. Avec Maven, vous pouvez créer, publier et déployer une tâche Flink avec Delta Lake sur Amazon. EMR

```
<project>
<properties>
  <scala.main.version>2.12</scala.main.version>
  <delta-connectors-version>0.6.0</delta-connectors-version>
  <flink-version>1.16.1</flink-version>
  <hadoop-version>3.1.0</hadoop-version>
</properties>

<dependencies>
  <dependency>
    <groupId>io.delta</groupId>
    <artifactId>delta-flink</artifactId>
    <version>${delta-connectors-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>io.delta</groupId>
    <artifactId>delta-standalone_${scala-main-version}</artifactId>
    <version>${delta-connectors-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-clients</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-parquet</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
  </dependency>
  <dependency>
    <groupId>org.apache.hadoop</groupId>
```



```

    <artifactId>hadoop-client</artifactId>
    <version>${hadoop-version}</version>
    <scope>provided</scope>
</dependency>
<dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-table-common</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
</dependency>
<dependency>
    <groupId>org.apache.flink</groupId>
    <artifactId>flink-table-runtime</artifactId>
    <version>${flink-version}</version>
    <scope>provided</scope>
</dependency>
</dependencies>

```

Écrire dans une table Delta avec Flink Datastream API

Utilisez l'exemple suivant pour créer un DeltaSink pour écrire dans la table avec un `deltaTablePath`:

```

public static DataStream<RowData> createDeltaSink(
    DataStream<RowData> stream,
    String deltaTablePath,
    RowType rowType) {
    Configuration configuration = new Configuration();
    DeltaSink<RowData> deltaSink = DeltaSink
        .forRowData(
            new org.apache.flink.core.fs.Path(deltaTablePath),
            configuration,
            rowType)
        .build();
    stream.sinkTo(deltaSink);
    return stream;
}

```

Lire à partir d'une table Delta avec Flink Datastream API

Utilisez l'exemple suivant pour créer un borné DeltaSource à lire dans le tableau avec un `deltaTablePath`:

```
public static DataStream<RowData> createBoundedDeltaSourceAllColumns(
    StreamExecutionEnvironment env,
    String deltaTablePath) {
    Configuration configuration = new Configuration();
    DeltaSource<RowData> deltaSource = DeltaSource
        .forBoundedRowData(
            new org.apache.flink.core.fs.Path(deltaTablePath),
            configuration)
        .build();

    return env.fromSource(deltaSource, WatermarkStrategy.noWatermarks(), "delta-
source");
}
```

Création de récepteurs avec prise en charge de plusieurs clusters pour le système autonome de Delta Lake

Utilisez l'exemple suivant pour créer une table DeltaSink à écrire avec un `deltaTablePath` [support multi-clusters](#) :

```
public DataStream<RowData> createDeltaSink(
    DataStream<RowData> stream,
    String deltaTablePath) {
    Configuration configuration = new Configuration();
    configuration.set("spark.delta.logStore.s3.impl",
        "io.delta.storage.S3DynamoDBLogStore");
    configuration.set("spark.io.delta.storage.S3DynamoDBLogStore.ddb.tableName",
        "delta_log");
    configuration.set("spark.io.delta.storage.S3DynamoDBLogStore.ddb.region", "us-
east-1");

    DeltaSink<RowData> deltaSink = DeltaSink
        .forRowData(
            new Path(deltaTablePath),
            configuration,
            rowType)
        .build();
    stream.sinkTo(deltaSink);
    return stream;
}
```

Exécuter la tâche Flink

Utilisez la commande suivante pour exécuter votre tâche :

```
flink run FlinkJob.jar
```

Utilisez un cluster Delta Lake avec Trino

Avec les EMR versions 6.9.0 et supérieures d'Amazon, vous pouvez utiliser Delta Lake avec votre cluster Trino.

Dans ce didacticiel, nous allons utiliser le cluster EMR Trino AWS CLI pour travailler avec Delta Lake sur Amazon.

Création d'un cluster Delta Lake

1. Créez un fichier, `delta_configurations.json`, et définissez des valeurs pour le catalogue de votre choix. Par exemple, si vous souhaitez utiliser le métastore Hive comme catalogue, le contenu de votre fichier doit être le suivant :

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore.uri":"thrift://localhost:9083"}}]
```

Si vous souhaitez utiliser le AWS Glue Catalog comme boutique, le contenu de votre fichier doit être le suivant :

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore":"glue"}}]
```

2. Créez un cluster avec la configuration suivante, en remplaçant **exemple Amazon S3 bucket path** et **subnet ID** par les vôtres.

```
aws emr create-cluster
  --release-label emr-6.9.0
  --applications Name=Trino
  --configurations file://delta_configurations.json
```

```
--region us-east-1 --name My_Spark_Delta_Cluster
--log-uri s3://DOC-EXAMPLE-BUCKET/
--instance-type m5.xlarge
--instance-count 2
--service-role EMR_DefaultRole_V2
--ec2-attributes
InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Initialisation de la session Trino pour Delta Lake

Pour initialiser la session Trino, exécutez la commande suivante

```
trino-cli --catalog delta
```

Écriture dans une table Delta Lake

Créez et écrivez dans votre table à l'aide des SQL commandes suivantes :

```
SHOW SCHEMAS;

CREATE TABLE default.delta_table (id int, data varchar, category varchar) WITH
( location = 's3://DOC-EXAMPLE-BUCKET/<prefix>');

INSERT INTO default.delta_table VALUES (1, 'a', 'c1'), (2, 'b', 'c2'), (3, 'c', 'c3');
```

Lecture à partir d'une table Delta Lake

Lisez le contenu de votre tableau à l'aide de la SQL commande suivante :

```
SELECT * from default.delta_table;
```

Utiliser un cluster Delta Lake avec Spark

À partir de EMR la version 6.9.0 d'Amazon, vous pouvez utiliser Delta Lake avec votre cluster Spark sans avoir besoin d'actions d'amorçage. Pour les EMR versions 6.8.0 et antérieures d'Amazon, vous pouvez utiliser les actions bootstrap pour préinstaller les dépendances nécessaires.

Les exemples suivants utilisent le AWS CLI pour travailler avec Delta Lake sur un cluster Amazon EMR Spark.

Pour utiliser Delta Lake sur Amazon EMR avec le AWS Command Line Interface, créez d'abord un cluster. Pour plus d'informations sur la façon de spécifier la classification de Delta Lake avec AWS Command Line Interface, voir [Fournir une configuration à l'aide du AWS Command Line Interface lorsque vous créez un cluster](#) ou [Fournir une configuration avec Java SDK lorsque vous créez un cluster](#).

1. Créez un fichier, `configurations.json`, contenant les éléments suivants :

```
[{"Classification":"delta-defaults", "Properties":{"delta.enabled":"true"} ]]
```

2. Créez un cluster avec la configuration suivante, en remplaçant l'exemple d'Amazon S3 **bucket path** et le **subnet ID** par le vôtre.

```
aws emr create-cluster
  --release-label emr-6.9.0
  --applications Name=Spark
  --configurations file://delta_configurations.json
  --region us-east-1
  --name My_Spark_Delta_Cluster
  --log-uri s3://DOC-EXAMPLE-BUCKET/
  --instance-type m5.xlarge
  --instance-count 2
  --service-role EMR_DefaultRole_V2
  --ec2-attributes
  InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Vous pouvez également créer un EMR cluster Amazon et une application Spark avec les fichiers suivants comme JAR dépendances dans une tâche Spark :

```
/usr/share/aws/delta/lib/delta-core.jar,
/usr/share/aws/delta/lib/delta-storage.jar,
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar
```

Note

Si vous utilisez les EMR versions 6.9.0 ou supérieures d'Amazon, utilisez à la `/usr/share/aws/delta/lib/delta-spark.jar` place de `/usr/share/aws/delta/lib/delta-core.jar`.

Pour plus d'informations, consultez [Soumission d'applications](#).

Pour inclure une dépendance JAR dans la tâche Spark, vous pouvez ajouter les propriétés de configuration suivantes à l'application Spark :

```
--conf "spark.jars=/usr/share/aws/delta/lib/delta-core.jar,  
/usr/share/aws/delta/lib/delta-storage.jar,  
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar"
```

Pour plus d'informations sur les dépendances entre les tâches de Spark, consultez [Gestion des dépendances](#).

Si vous utilisez les EMR versions 6.9.0 ou supérieures d'Amazon, ajoutez plutôt la `/usr/share/aws/delta/lib/delta-spark.jar` configuration.

```
--conf "spark.jars=/usr/share/aws/delta/lib/delta-spark.jar,  
/usr/share/aws/delta/lib/delta-storage.jar,  
/usr/share/aws/delta/lib/delta-storage-s3-dynamodb.jar"
```

Initialisation d'une session Spark pour Delta Lake

Les exemples suivants montrent comment lancer le shell interactif Spark, utiliser Spark submit ou utiliser Amazon EMR Notebooks pour travailler avec Delta Lake sur AmazonEMR.

spark-shell

1. Connectez-vous au nœud principal à l'aide deSSH. Pour plus d'informations, consultez [Se connecter au nœud principal SSH à l'aide](#) de l'Amazon EMR Management Guide.
2. Entrez la commande suivante pour lancer le shell Spark. Pour utiliser la PySpark coque, remplacez-la `spark-shell` par `parpyspark`.

```
spark-shell \
  --conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \
  --conf
"spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

Si vous exécutez les EMR versions 6.15.0 ou supérieures d'Amazon, vous devez également utiliser les configurations suivantes pour utiliser un contrôle d'accès précis basé sur Lake Formation with Delta Lake.

```
spark-shell \
  --conf
spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordser
\
  --conf
spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog \

  --conf spark.sql.catalog.spark_catalog.lf.managed=true
```

spark-submit

1. Connectez-vous au nœud principal à l'aide de SSH. Pour plus d'informations, consultez [Se connecter au nœud principal SSH à l'aide](#) de l'Amazon EMR Management Guide.
2. Entrez la commande suivante pour lancer la session Spark pour Delta Lake.

```
spark-submit
--conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension"
--conf
"spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

Si vous exécutez les EMR versions 6.15.0 ou supérieures d'Amazon, vous devez également utiliser les configurations suivantes pour utiliser un contrôle d'accès précis basé sur Lake Formation with Delta Lake.

```
spark-submit \ `
  --conf
spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordser
```

```
--conf
  spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog \

--conf spark.sql.catalog.spark_catalog.lf.managed=true
```

EMR Studio notebooks

Pour initialiser une session Spark à l'aide des blocs-notes Amazon EMR Studio, configurez votre session Spark à l'aide de %%configure la commande magique dans votre EMR bloc-notes Amazon, comme dans l'exemple suivant. Pour plus d'informations, consultez la section [Utiliser la magie EMR des blocs-notes dans](#) le guide de EMRgestion Amazon.

```
%%configure -f
{
  "conf": {
    "spark.sql.extensions": "io.delta.sql.DeltaSparkSessionExtension",
    "spark.sql.catalog.spark_catalog":
"org.apache.spark.sql.delta.catalog.DeltaCatalog"
  }
}
```

Si vous exécutez les EMR versions 6.15.0 ou supérieures d'Amazon, vous devez également utiliser les configurations suivantes pour utiliser un contrôle d'accès précis basé sur Lake Formation with Delta Lake.

```
%%configure -f
{
  "conf": {
    "spark.sql.extensions":
"io.delta.sql.DeltaSparkSessionExtension,com.amazonaws.emr.recordserver.connector.spark.sql"
    "spark.sql.catalog.spark_catalog":
"org.apache.spark.sql.delta.catalog.DeltaCatalog",
    "spark.sql.catalog.spark_catalog.lf.managed": "true"
  }
}
```

Écriture dans une table Delta Lake

L'exemple suivant montre comment créer un jeu de données Delta Lake DataFrame et l'écrire sous forme de jeu de données. L'exemple montre comment travailler avec des ensembles de données

avec le shell Spark lorsque vous êtes connecté au nœud principal en SSH tant qu'utilisateur hadoop par défaut.

Note

Pour coller des exemples de code dans le shell Spark, tapez `:paste` à l'invite, collez l'exemple, puis appuyez sur CTRL + D.

PySpark

Spark inclut un shell basé sur Python, `pyspark`, que vous pouvez utiliser pour créer un prototype de programmes Spark écrits en Python. Tout comme pour `spark-shell`, invoquez `pyspark` sur le nœud primaire.

```
## Create a DataFrame
data = spark.createDataFrame([("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
                              ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
                              ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
                              ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")],
                              ["id", "creation_date", "last_update_time"])

## Write a DataFrame as a Delta Lake dataset to the S3 location
spark.sql("""CREATE TABLE IF NOT EXISTS delta_table (id string, creation_date
string,
last_update_time string)
USING delta location
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table'""");

data.writeTo("delta_table").append()
```

Scala

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._

// Create a DataFrame
val data = Seq(("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
              ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
              ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
              ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")).toDF("id", "creation_date",
"last_update_time")
```

```
// Write a DataFrame as a Delta Lake dataset to the S3 location
spark.sql("""CREATE TABLE IF NOT EXISTS delta_table (id string,
creation_date string,
last_update_time string)
USING delta location
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table'""");

data.write.format("delta").mode("append").saveAsTable("delta_table")
```

SQL

```
-- Create a Delta Lake table with the S3 location
CREATE TABLE delta_table(id string,
creation_date string,
last_update_time string)
USING delta LOCATION
's3://DOC-EXAMPLE-BUCKET/example-prefix/db/delta_table';

-- insert data into the table
INSERT INTO delta_table VALUES ("100", "2015-01-01",
"2015-01-01T13:51:39.340396Z"),
("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
("103", "2015-01-01", "2015-01-01T13:51:40.519832Z");
```

Lecture à partir d'une table Delta Lake

PySpark

```
ddf = spark.table("delta_table")
ddf.show()
```

Scala

```
val ddf = spark.table("delta_table")
ddf.show()
```

SQL

```
SELECT * FROM delta_table;
```

Utilisez un cluster Delta Lake avec Spark et AWS Glue

Pour utiliser le catalogue AWS Glue comme métastore pour les tables Delta Lake, créez un cluster en suivant les étapes suivantes. Pour plus d'informations sur la spécification de la classification de Delta Lake à l'aide de AWS Command Line Interface, voir [Fournir une configuration à l'aide du AWS Command Line Interface lorsque vous créez un cluster](#) ou [Fournissez une configuration à l'aide de Java SDK lorsque vous créez un cluster](#).

Création d'un cluster Delta Lake

1. Créez un fichier, `configurations.json`, contenant les éléments suivants :

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
{"Classification":"spark-hive-site",
  "Properties":
  {"hive.metastore.client.factory.class":"com.amazonaws.glue.catalog.metastore.AWSGlueDataCat
```

2. Créez un cluster avec la configuration suivante, en remplaçant **exemple Amazon S3 bucket path** et **subnet ID** par les vôtres.

```
aws emr create-cluster
  --release-label emr-6.9.0
  --applications Name=Spark
  --configurations file://delta_configurations.json
  --region us-east-1
  --name My_Spark_Delta_Cluster
  --log-uri s3://DOC-EXAMPLE-BUCKET/
  --instance-type m5.xlarge
  --instance-count 2
  --service-role EMR_DefaultRole_V2
  --ec2-attributes
  InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Considérations et restrictions

- Delta Lake est compatible avec les EMR versions 6.9.0 et supérieures d'Amazon. Vous pouvez utiliser [Apache Spark](#) 3.x sur des EMR clusters Amazon avec des tables Delta.

- Nous vous recommandons d'utiliser le URI schéma s3 pour les chemins de localisation S3 au lieu de s3a pour des performances, une sécurité et une fiabilité optimales. Pour plus d'informations, consultez [Utilisation des systèmes de stockage et de fichiers](#).
- Avec Amazon EMR 7.0, Delta Universal Format (UniForm) les instructions Convert-to-Iceberg ne sont pas prises en charge.
- Avec Amazon EMR 6.9 et 6.10, lorsque vous stockez des données de table Delta Lake dans Amazon S3, les données de colonne deviennent des données NULL après l'opération de renommage des colonnes. Ce problème est résolu avec Amazon EMR 6.11. Pour plus d'informations sur l'opération expérimentale de changement de nom de colonne, consultez [Opération de renommage de colonne](#) dans le Guide de l'utilisateur de Delta Lake.
- Si vous créez une base de données dans le catalogue de données AWS Glue en dehors d'Apache Spark, le LOCATION champ de la base de données peut être vide. Dans la mesure où Spark n'autorise pas la création de bases de données avec une propriété d'emplacement vide, vous obtiendrez l'erreur suivante si vous utilisez Spark dans Amazon EMR pour créer une table Delta dans une base de données Glue et que la base de données possède une LOCATION propriété vide :

```
IllegalArgumentException: Can not create a Path from an empty string
```

Pour résoudre ce problème, créez la base de données dans le catalogue de données avec un chemin valide et non vide pour le champ LOCATION. Pour connaître les étapes de mise en œuvre de cette solution, consultez la section [Exception d'argument illégal lors de la création d'une table](#) dans le Guide de l'utilisateur Amazon Athena.

Historique des versions de Delta

Le tableau suivant répertorie la version de Delta incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version Delta

Étiquette Amazon EMR Release	Version Delta	Composants installés avec Delta
emr-7.2.0	3.1.0	Not available.
emr-7.1.0	3.0.0	Not available.
emr-7.0.0	3.0.0	Not available.
emr-6.15.0	2.4.0	Not available.
emr-6.14.0	2.4.0	Not available.
emr-6.13.0	2.4.0	Not available.
emr-6.12.0	2.4.0	Not available.
emr-6.11.1	2.2.0	Not available.
emr-6.11.0	2.2.0	Not available.
emr-6.10.1	2.2.0	Not available.
emr-6.10.0	2.2.0	Not available.
emr-6.9.1	2.1.0	Not available.
emr-6.9.0	2.1.0	Not available.

Apache Flink

[Apache Flink](#) est un moteur de flux de données en streaming permettant de traiter des flux en temps réel sur des sources de données à haut débit. Flink prend en charge la sémantique de l'heure des out-of-order événements pour les événements, la sémantique d'une seule fois, le contrôle de la contre-pression, et est APIs optimisé pour écrire à la fois des applications de streaming et des applications par lots.

En outre, Flink a des connecteurs pour les sources de données tierces, telles que :

- [Amazon Kinesis Data Streams](#)
- [Apache Kafka](#)
- [Connecteur Flink Elasticsearch](#)
- [Diffusion sur Twitter API](#)
- [Cassandra](#)

Amazon EMR prend en charge Flink en tant qu'YARNApplication afin que vous puissiez gérer les ressources avec d'autres applications au sein d'un cluster. Flink-on- vous YARN permet de soumettre des tâches Flink transitoires ou de créer un cluster de longue durée qui accepte plusieurs tâches et alloue des ressources en fonction de la réservation globale. YARN

Flink est inclus dans les EMR versions 5.1.0 et ultérieures d'Amazon.

Note

Support pour cette `FlinkKinesisConsumer` classe a été ajouté dans la EMR version 5.2.1 d'Amazon.

Le tableau suivant répertorie la version de Flink incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Flink.

Pour la version des composants installés avec Flink dans cette version, voir [Versions des composants de la version 7.2.0](#).

Informations sur la version de Flink pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Flink	Composants installés avec Flink
emr-7.2.0	Flink 1.18.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta

Le tableau suivant répertorie la version de Flink incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Flink.

Pour connaître la version des composants installés avec Flink dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Flink pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Flink	Composants installés avec Flink
emr-6.15.0	Flink 1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client,

Étiquette EMR de sortie Amazon	Version Flink	Composants installés avec Flink
		flink-jobmanager-config, hudi, delta-standalone-connectors

Le tableau suivant répertorie la version de Flink incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Flink.

Pour la version des composants installés avec Flink dans cette version, voir la [version 5.36.2](#) Versions des composants.

Informations de version de Flink pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Flink	Composants installés avec Flink
emr-5.36.2	Flink 1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Rubriques

- [Création d'un cluster avec Flink](#)
- [Configuration de Flink sur Amazon EMR](#)
- [Utilisation des tâches Flink dans Amazon EMR](#)
- [Utilisation du shell Scala](#)
- [Recherche de l'interface web Flink](#)
- [Outil de mise à l'échelle automatique Flink](#)

- [Optimisation des temps de redémarrage pour les opérations de récupération et de mise à l'échelle de tâches](#)
- [Travailler avec les jobs Flink de Zeppelin dans Amazon EMR](#)
- [Historique des publications de Flink](#)

Création d'un cluster avec Flink

Vous pouvez lancer un cluster avec le AWS Management Console AWS CLI, ou un AWS SDK.

Pour lancer un cluster avec Flink installé depuis la console

1. Ouvrez la EMR console Amazon à l'[adresse https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr).
2. Choisissez Créer un cluster et Go to advanced options (Aller aux options avancées).
3. Pour la configuration logicielle, choisissez la EMRversion emr-5.1.0 ou ultérieure.
4. Choisissez Flink comme application, de même que toute autre application à installer.
5. Sélectionnez d'autres options si nécessaire, puis choisissez Create cluster (Créer un cluster).

Pour lancer un cluster avec Flink depuis AWS CLI

- Créez le cluster à l'aide de la commande suivante :

```
aws emr create-cluster --release-label emr-7.2.0 \  
--applications Name=Flink \  
--configurations file://./configurations.json \  
--region us-east-1 \  
--log-uri s3://myLogUri \  
--instance-type m5.xlarge \  
--instance-count 2 \  
--service-role EMR_DefaultRole_V2 \  
--ec2-attributes KeyName=MyKeyName,InstanceProfile=EMR_EC2_DefaultRole \  
--steps Type=CUSTOM_JAR,Jar=command-runner.jar,Name=Flink_Long_Running_Session,\  
Args=flink-yarn-session,-d
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

Configuration de Flink sur Amazon EMR

Configuration de Flink avec Hive Metastore et Glue Catalog

Les EMR versions 6.9.0 et supérieures d'Amazon prennent en charge à la fois Hive Metastore et AWS Glue Catalog avec le connecteur Apache Flink vers Hive. Cette section décrit les étapes nécessaires pour configurer [AWS Glue Catalog](#) et [Hive Metastore](#) avec Flink.

Rubriques

- [Utilisation de Hive Metastore](#)
- [Utiliser le catalogue AWS de données Glue](#)

Utilisation de Hive Metastore

1. Créez un EMR cluster avec la version 6.9.0 ou supérieure et au moins deux applications : Hive et Flink.
2. Utilisez [script runner](#) pour exécuter le script suivant en tant que fonction d'étape :

```
hive-metastore-setup.sh
```

```
sudo cp /usr/lib/hive/lib/antlr-runtime-3.5.2.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/hive-exec-3.1.3*.jar /lib/flink/lib
sudo cp /usr/lib/hive/lib/libfb303-0.9.3.jar /lib/flink/lib
sudo cp /usr/lib/flink/opt/flink-connector-hive_2.12-1.15.2.jar /lib/flink/lib
sudo chmod 755 /usr/lib/flink/lib/antlr-runtime-3.5.2.jar
sudo chmod 755 /usr/lib/flink/lib/hive-exec-3.1.3*.jar
sudo chmod 755 /usr/lib/flink/lib/libfb303-0.9.3.jar
sudo chmod 755 /usr/lib/flink/lib/flink-connector-hive_2.12-1.15.2.jar
```

Add step ✕

Step type Custom JAR

Name*

JAR location* JAR location maybe a path into S3 or a fully qualified java class in the classpath.

Arguments These are passed to the main function in the JAR. If the JAR does not specify a main class in its manifest file you can specify another class name as the first argument.

Action on failure What happens if the step fails

Utiliser le catalogue AWS de données Glue

1. Créez un EMR cluster avec la version 6.9.0 ou supérieure et au moins deux applications : Hive et Flink.
2. Sélectionnez Utiliser les métadonnées de la table Hive dans les paramètres du catalogue de données AWS Glue pour activer le catalogue de données dans le cluster.
3. Utilisez [le script Runner](#) pour exécuter le script suivant en tant que fonction d'étape : [Exécutez des commandes et des scripts sur un EMR cluster Amazon](#) :

glue-catalog-setup.sh

```

sudo cp /usr/lib/hive/auxlib/aws-glue-datacatalog-hive3-client.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/antlr-runtime-3.5.2.jar /usr/lib/flink/lib
sudo cp /usr/lib/hive/lib/hive-exec-3.1.3*.jar /lib/flink/lib
sudo cp /usr/lib/hive/lib/libfb303-0.9.3.jar /lib/flink/lib
sudo cp /usr/lib/flink/opt/flink-connector-hive_2.12-1.15.2.jar /lib/flink/lib
sudo chmod 755 /usr/lib/flink/lib/aws-glue-datacatalog-hive3-client.jar
sudo chmod 755 /usr/lib/flink/lib/antlr-runtime-3.5.2.jar
sudo chmod 755 /usr/lib/flink/lib/hive-exec-3.1.3*.jar
sudo chmod 755 /usr/lib/flink/lib/libfb303-0.9.3.jar
sudo chmod 755 /usr/lib/flink/lib/flink-connector-hive_2.12-1.15.2.jar

```

Add step ✕

Step type Custom JAR

Name*

JAR location* JAR location maybe a path into S3 or a fully qualified java class in the classpath.

Arguments These are passed to the main function in the JAR. If the JAR does not specify a main class in its manifest file you can specify another class name as the first argument.

Action on failure What happens if the step fails

Cancel
Save

Configurer Flink avec un fichier de configuration

Vous pouvez utiliser la EMR configuration Amazon API pour configurer Flink à l'aide d'un fichier de configuration. Les fichiers configurables dans le API sont les suivants :

- `flink-conf.yaml`
- `log4j.properties`
- `flink-log4j-session`
- `log4j-cli.properties`

Le fichier de configuration principal de Flink est `flink-conf.yaml`.

Pour configurer le nombre d'emplacements de tâches utilisés par Flink à partir du fichier AWS CLI

1. Créez un fichier, `configurations.json`, contenant les éléments suivants :

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "taskmanager.numberOfTaskSlots": "2"
    }
  }
]
```

]

2. Créez ensuite un cluster à l'aide de la configuration suivante :

```
aws emr create-cluster --release-label emr-7.2.0 \  
--applications Name=Flink \  
--configurations file:///./configurations.json \  
--region us-east-1 \  
--log-uri s3:///myLogUri \  
--instance-type m5.xlarge \  
--instance-count 2 \  
--service-role EMR_DefaultRole_V2 \  
--ec2-attributes KeyName=YourKeyName,InstanceProfile=EMR_EC2_DefaultRole
```

Note

Vous pouvez également modifier certaines configurations avec le FlinkAPI. Pour plus d'informations, consultez [Concepts](#) dans la documentation Flink.

Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier des classifications de configuration supplémentaires pour chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, utilisez la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).

Options de parallélisme

En tant que propriétaire de votre application, c'est vous qui savez le mieux quelles ressources attribuer aux tâches dans Flink. Pour les exemples de cette documentation, utilisez le même nombre de tâches que les instances de tâches que vous utilisez pour l'application. Généralement, nous recommandons cela pour le niveau initial de parallélisme, mais vous pouvez également augmenter la granularité de parallélisme à l'aide des emplacements de tâches, qui ne doivent généralement pas dépasser le nombre de [cœurs virtuels](#) par instance. Pour plus d'informations sur l'architecture de Flink, consultez [Concepts](#) dans la documentation Flink.

Configuration de Flink sur un EMR cluster comportant plusieurs nœuds principaux

Le JobManager de Flink reste disponible pendant le processus de basculement du nœud principal dans un EMR cluster Amazon comportant plusieurs nœuds principaux. À partir d'Amazon EMR 5.28.0, la JobManager haute disponibilité est également activée automatiquement. Aucune configuration manuelle n'est nécessaire.

Avec EMR les versions 5.27.0 ou antérieures d'Amazon, il JobManager s'agit d'un point de défaillance unique. En cas d' JobManager échec, il perd tous les états des tâches et ne reprend pas les tâches en cours d'exécution. Vous pouvez activer la JobManager haute disponibilité en configurant le nombre de tentatives d'application, le point de contrôle et ZooKeeper en activant le stockage d'état pour Flink, comme le montre l'exemple suivant :

```
[
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.resourcemanager.am.max-attempts": "10"
    }
  },
  {
    "Classification": "flink-conf",
    "Properties": {
      "yarn.application-attempts": "10",
      "high-availability": "zookeeper",
      "high-availability.zookeeper.quorum": "%{hiera('hadoop:zk')}",
      "high-availability.storageDir": "hdfs:///user/flink/recovery",
      "high-availability.zookeeper.path.root": "/flink"
    }
  }
]
```

Vous devez configurer le nombre maximal de tentatives de master d'application pour Flink YARN et de tentatives d'application pour Flink. Pour plus d'informations, consultez [la section Configuration de la haute disponibilité du YARN cluster](#). Vous pouvez également configurer le point de contrôle Flink pour que les tâches de JobManager restauration en cours d'exécution redémarrées soient restaurées à partir de points de contrôle précédemment terminés. Pour plus d'informations, consultez [Flink checkpointing](#).

Configuration de la taille du processus de mémoire

Pour EMR les versions d'Amazon qui utilisent Flink 1.11.x, vous devez configurer la taille totale du processus de mémoire pour JobManager (`jobmanager.memory.process.size`) et TaskManager (`taskmanager.memory.process.size`) dans `flink-conf.yaml`. Vous pouvez définir ces valeurs soit en configurant le cluster avec la configuration, API soit en décommentant manuellement ces champs via SSH. Flink fournit les valeurs par défaut suivantes.

- `jobmanager.memory.process.size` : 1600m
- `taskmanager.memory.process.size` : 1728m

Pour exclure le JVM métaspace et la surcharge, utilisez la taille de mémoire totale de Flink (`taskmanager.memory.flink.size`) au lieu de `taskmanager.memory.process.size`. La valeur par défaut du paramètre `taskmanager.memory.process.size` est 1280m. Il n'est pas recommandé de définir à la fois `taskmanager.memory.process.size` et `taskmanager.memory.flink.size`.

Toutes les EMR versions d'Amazon qui utilisent Flink 1.12.0 et versions ultérieures ont les valeurs par défaut répertoriées dans l'ensemble open source pour Flink comme valeurs par défaut sur AmazonEMR. Vous n'avez donc pas besoin de les configurer vous-même.

Configuration de la taille du fichier de sortie du journal

Les conteneurs d'applications Flink créent et écrivent dans trois types de fichiers journaux : fichiers `.out`, fichiers `.log` et fichiers `.err`. Seuls `.err` les fichiers sont compressés et supprimés du système de fichiers, tandis que les fichiers journaux `.log` et `.out` restent dans le système de fichiers. Pour garantir la gérabilité de ces fichiers de sortie et la stabilité du cluster, vous pouvez configurer la rotation des journaux dans `log4j.properties` afin de définir un nombre maximum de fichiers et de limiter leur taille.

Amazon EMR versions 5.30.0 et ultérieures

À partir d'Amazon EMR 5.30.0, Flink utilise le framework de journalisation `log4j2` avec le nom de classification de configuration. L'exemple de configuration suivant illustre `flink-log4j`. le format `log4j2`.

```
[  
  {
```

```

"Classification": "flink-log4j",
"Properties": {
  "appender.main.name": "MainAppender",
  "appender.main.type": "RollingFile",
  "appender.main.append" : "false",
  "appender.main.fileName" : "${sys:log.file}",
  "appender.main.filePattern" : "${sys:log.file}.%i",
  "appender.main.layout.type" : "PatternLayout",
  "appender.main.layout.pattern" : "%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p %-60c %x - %m
%n",
  "appender.main.policies.type" : "Policies",
  "appender.main.policies.size.type" : "SizeBasedTriggeringPolicy",
  "appender.main.policies.size.size" : "100MB",
  "appender.main.strategy.type" : "DefaultRolloverStrategy",
  "appender.main.strategy.max" : "10"
},
}
]

```

Amazon EMR versions 5.29.0 et antérieures

Avec EMR les versions 5.29.0 et antérieures d'Amazon, Flink utilise le framework de journalisation log4j. L'exemple de configuration suivant illustre le format log4j.

```

[
{
  "Classification": "flink-log4j",
  "Properties": {
    "log4j.appender.file": "org.apache.log4j.RollingFileAppender",
    "log4j.appender.file.append": "true",
    # keep up to 4 files and each file size is limited to 100MB
    "log4j.appender.file.MaxFileSize": "100MB",
    "log4j.appender.file.MaxBackupIndex": 4,
    "log4j.appender.file.layout": "org.apache.log4j.PatternLayout",
    "log4j.appender.file.layout.ConversionPattern": "%d{yyyy-MM-dd HH:mm:ss,SSS} %-5p
%-60c %x - %m%n"
  },
}
]

```


Configurer Flink pour qu'il fonctionne avec Java 11

EMR Les versions 6.12.0 et supérieures d'Amazon fournissent un support d'exécution Java 11 pour Flink. Les sections suivantes décrivent comment configurer le cluster pour fournir un support d'exécution Java 11 pour Flink.

Rubriques

- [Configurer Flink pour Java 11 lorsque vous créez un cluster](#)
- [Configurer Flink pour Java 11 sur un cluster en cours d'exécution](#)
- [Confirmez le runtime Java pour Flink sur un cluster en cours d'exécution](#)

Configurer Flink pour Java 11 lorsque vous créez un cluster

Procédez comme suit pour créer un EMR cluster avec Flink et Java 11 Runtime. Le fichier de configuration dans lequel vous ajoutez la prise en charge de Java 11 est `flink-conf.yaml`.

Console

Pour créer un cluster avec Flink et Java 11 Runtime dans la console

1. Connectez-vous à la AWS Management Console EMR console Amazon et ouvrez-la à l'adresse <https://console.aws.amazon.com/emr>.
2. Choisissez Clusters sous EMR activé EC2 dans le volet de navigation, puis Create cluster.
3. Sélectionnez Amazon EMR version 6.12.0 ou supérieure, puis choisissez d'installer l'application Flink. Sélectionnez les autres applications que vous souhaitez installer sur votre cluster.
4. Poursuivez la configuration de votre cluster. Dans la section facultatifs Paramètres logiciels, utilisez l'option par défaut Entrer un configuration et entrez la configuration suivante :

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "containerized.taskmanager.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
      "containerized.master.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
      "env.java.home": "/usr/lib/jvm/jre-11"
    }
  }
]
```

```
]
```

5. Continuez à configurer et à lancer votre cluster.

AWS CLI

Pour créer un cluster avec Flink et le runtime Java 11 à partir du CLI

1. Créez un fichier de configuration `configurations.json` qui configure Flink pour utiliser Java 11.

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "containerized.taskmanager.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "containerized.master.env.JAVA_HOME":"/usr/lib/jvm/jre-11",
      "env.java.home":"/usr/lib/jvm/jre-11"
    }
  }
]
```

2. À partir de AWS CLI, créez un nouveau EMR cluster avec Amazon EMR version 6.12.0 ou supérieure, et installez l'application Flink, comme indiqué dans l'exemple suivant :

```
aws emr create-cluster --release-label emr-6.12.0 \
--applications Name=Flink \
--configurations file://./configurations.json \
--region us-east-1 \
--log-uri s3://myLogUri \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes KeyName=YourKeyName,InstanceProfile=EMR_EC2_DefaultRole
```

Configurer Flink pour Java 11 sur un cluster en cours d'exécution

Procédez comme suit pour mettre à jour un EMR cluster en cours d'exécution avec Flink et Java 11 Runtime. Le fichier de configuration dans lequel vous ajoutez la prise en charge de Java 11 est `flink-conf.yaml`.

Console

Pour mettre à jour un cluster en cours d'exécution avec Flink et Java 11 Runtime dans la console

1. Connectez-vous à la AWS Management Console EMR console Amazon et ouvrez-la à l'adresse <https://console.aws.amazon.com/emr>.
2. Choisissez Clusters sous EMR activé EC2 dans le volet de navigation, puis sélectionnez le cluster que vous souhaitez mettre à jour.

Note

Le cluster doit utiliser Amazon EMR version 6.12.0 ou ultérieure pour prendre en charge Java 11.

3. Sélectionnez l'onglet Configurations.
4. Dans la section Configurations du groupe d'instances, sélectionnez le groupe d'instances en cours d'exécution que vous souhaitez mettre à jour, puis choisissez Reconfigurer dans le menu d'actions de la liste.
5. Reconfigurez le groupe d'instances avec l'option Modifier les attributs comme suit. Sélectionnez Ajouter une nouvelle configuration après chacune d'entre elles.

Classification	Propriété	Valeur
flink-conf	containerized.taskmanager.env.JAVA_HOME	/usr/lib/jvm/jre-11
flink-conf	containerized.master.env.JAVA_HOME	/usr/lib/jvm/jre-11
flink-conf	env.java.home	/usr/lib/jvm/jre-11

6. Sélectionnez Enregistrer les modifications pour ajouter les configurations.

AWS CLI

Pour mettre à jour un cluster en cours d'exécution afin d'utiliser Flink et le runtime Java 11 à partir du CLI

Utilisez la commande `modify-instance-groups` pour spécifier une nouvelle configuration pour un groupe d'instances dans un cluster en cours d'exécution.

1. Créez d'abord un fichier de configuration `configurations.json` qui configure Flink pour utiliser Java 11. Dans l'exemple suivant, remplacez `ig-1xxxxxxx9` avec l'ID du groupe d'instances que vous souhaitez reconfigurer. Enregistrez le fichier dans le même répertoire que celui où vous exécuterez la commande `modify-instance-groups`.

```
[
  {
    "InstanceGroupId": "ig-1xxxxxxx9",
    "Configurations": [
      {
        "Classification": "flink-conf",
        "Properties": {
          "containerized.taskmanager.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
          "containerized.master.env.JAVA_HOME": "/usr/lib/jvm/jre-11",
          "env.java.home": "/usr/lib/jvm/jre-11"
        },
        "Configurations": []
      }
    ]
  }
]
```

2. À partir de AWS CLI, exécutez la commande suivante. Remplacez l'ID du groupe d'instances que vous souhaitez reconfigurer :

```
aws emr modify-instance-groups --cluster-id j-2AL4XXXXXX5T9 \
--instance-groups file://configurations.json
```

Confirmez le runtime Java pour Flink sur un cluster en cours d'exécution

Pour déterminer l'environnement d'exécution Java d'un cluster en cours d'exécution, connectez-vous au nœud principal avec, SSH comme décrit dans la section [Se connecter au nœud principal avec SSH](#). Ensuite, exécutez la commande suivante :

```
ps -ef | grep flink
```

La commande `ps` associée à l'option `-ef` répertorie tous les processus en cours d'exécution sur le système. Vous pouvez filtrer cette sortie avec `grep` pour trouver les mentions de la chaîne `flink`. Vérifiez le résultat pour la valeur Java Runtime Environment (JRE), `jre-XX`. Dans le résultat suivant, `jre-11` indique que Java 11 est activé lors de l'exécution de Flink.

```
flink    19130      1  0 09:17 ?          00:00:15 /usr/lib/jvm/jre-11/bin/
java -Djava.io.tmpdir=/mnt/tmp -Dlog.file=/usr/lib/flink/log/flink-flink-
historyserver-0-ip-172-31-32-127.log -Dlog4j.configuration=file:/usr/lib/flink/conf/
log4j.properties -Dlog4j.configurationFile=file:/usr/lib/flink/conf/log4j.properties
-Dlogback.configurationFile=file:/usr/lib/flink/conf/logback.xml -classpath /usr/lib/
flink/lib/flink-cep-1.17.0.jar:/usr/lib/flink/lib/flink-connector-files-1.17.0.jar:/
usr/lib/flink/lib/flink-csv-1.17.0.jar:/usr/lib/flink/lib/flink-json-1.17.0.jar:/usr/
lib/flink/lib/flink-scala_2.12-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-java-
uber-1.17.0.jar:/usr/lib/flink/lib/flink-table-api-scala-bridge_2.12-1.17.0.
```

Vous pouvez également [vous connecter au nœud principal avec](#) une commande SSH et démarrer une YARN session Flink avec cette commande `flink-yarn-session -d`. La sortie montre la machine virtuelle Java (JVM) pour Flink, `java-11-amazon-corretto` dans l'exemple suivant :

```
2023-05-29 10:38:14,129 INFO  org.apache.flink.configuration.GlobalConfiguration
[] - Loading configuration property: containerized.master.env.JAVA_HOME, /usr/lib/
jvm/java-11-amazon-corretto.x86_64
```

Utilisation des tâches Flink dans Amazon EMR

Il existe plusieurs manières d'interagir avec Flink sur Amazon EMR : via la console, via l'interface Flink située dans l'interface utilisateur de suivi et via ResourceManager la ligne de commande. Vous pouvez envoyer un fichier JAR à une application Flink avec n'importe lequel de ces outils. Une fois soumis un fichier JAR, il devient une tâche gérée par le Flink JobManager. JobManager Il se trouve sur le nœud YARN qui héberge le daemon Application Master de la session Flink.

Vous pouvez exécuter une application Flink en tant que tâche YARN sur un cluster à exécution longue ou sur un cluster transitoire. Sur un cluster à exécution longue, vous pouvez soumettre plusieurs tâches Flink à un cluster Flink s'exécutant sur Amazon EMR. Si vous exécutez une tâche Flink sur un cluster transitoire, votre cluster Amazon EMR n'existe que le temps de l'exécution de l'application Flink, et vous n'êtes donc facturé que pour les ressources et le temps utilisés. Vous pouvez soumettre une tâche Flink avec l'opération d'API Amazon EMR `AddSteps`, en tant qu'argument d'étape de l'opération `RunJobFlow` et via les commandes AWS CLI `add-steps` ou `create-cluster`.

Démarrez une application Flink YARN en tant qu'étape sur un cluster de longue durée

Pour démarrer une application Flink à laquelle plusieurs clients peuvent soumettre du travail via les opérations de l'API YARN, vous devez soit créer un cluster, soit ajouter une application Flink à un cluster existant. Pour obtenir des instructions sur la création d'un cluster, veuillez consulter [Création d'un cluster avec Flink](#). Pour démarrer une session YARN sur un cluster existant, procédez comme suit à partir de la console, de l'AWS CLI ou du kit SDK Java.

Note

La commande `flink-yarn-session` a été ajoutée dans la version 5.5.0 d'Amazon EMR comme habillage pour le script `yarn-session.sh`, afin de simplifier l'exécution. Si vous utilisez une version antérieure d'Amazon EMR, remplacez `bash -c "/usr/lib/flink/bin/yarn-session.sh -d"` par Arguments dans la console ou par Args. dans la commande AWS CLI.

Pour soumettre une tâche Flink sur un cluster existant depuis la console

Soumettez la session Flink avec la commande `flink-yarn-session` dans un cluster existant.

1. [Ouvrez la console Amazon EMR à l'adresse `https://console.aws.amazon.com/emr`](https://console.aws.amazon.com/emr).
2. Dans la liste des clusters, sélectionnez celui que vous avez lancé précédemment.
3. Dans la page des détails de cluster, sélectionnez Steps (Étapes), Add Step (Ajouter une étape).
4. Utilisez les directives suivantes pour saisir les paramètres, puis sélectionnez Ajouter.

Paramètre	Description
Type d'étape	JAR personnalisé
Nom	Nom permettant d'identifier la nouvelle étape. Par exemple, < <i>example-flink-step-name</i> >.
Emplacement JAR	command-runner.jar
Arguments	La commande <code>flink-yarn-session</code> avec les arguments appropriés pour votre application. Par exemple, <code>flink-yarn-session -d</code> démarre une session Flink au sein de votre cluster YARN dans un état détaché (-d). Consultez Configuration de YARN dans la dernière documentation Flink pour plus de détails sur les arguments.

Pour soumettre une tâche Flink sur un cluster existant à l'aide de l'AWS CLI

- Utilisez la commande `add-steps` pour ajouter une tâche Flink à un cluster de longue durée. L'exemple de commande suivant indique `Args="flink-yarn-session", "-d"` pour démarrer une session Flink au sein de votre cluster YARN dans un état détaché (-d). Consultez [Configuration de YARN](#) dans la dernière documentation Flink pour plus de détails sur les arguments.

```
aws emr add-steps --cluster-id <j-XXXXXXXX> --steps Type=CUSTOM_JAR,Name=<example-flink-step-name>,Jar=command-runner.jar,Args="flink-yarn-session","-d"
```

Soumettre du travail à une application Flink existante sur un cluster de longue durée

Si vous possédez déjà une application Flink sur un cluster de longue durée, vous pouvez spécifier l'ID de l'application Flink du cluster afin d'y soumettre du travail. Pour obtenir l'ID de l'application, `yarn application -list` exécutez l'opération [YarnClient](#) API AWS CLI ou via l'API :

```
$ yarn application -list
16/09/07 19:32:13 INFO client.RMProxy: Connecting to ResourceManager at
ip-10-181-83-19.ec2.internal/10.181.83.19:8032
Total number of applications (application-types: [] and states: [SUBMITTED, ACCEPTED,
RUNNING]):1
Application-Id      Application-Name      Application-Type      User      Queue      State
Final-State      Progress      Tracking-URL
application_1473169569237_0002      Flink session with 14 TaskManagers (detached)
Apache Flink      hadoop      default      RUNNING      UNDEFINED
100% http://ip-10-136-154-194.ec2.internal:33089
```

L'identifiant de l'application pour cette session Flink est `application_1473169569237_0002`, que vous pouvez utiliser pour soumettre du travail à l'application à partir de l'AWS CLI ou d'un kit SDK.

Exemple SDK pour Java

```
List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

HadoopJarStepConfig flinkWordCountConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("flink", "run", "-m", "yarn-cluster", "-yid",
"application_1473169569237_0002", "-yn", "2", "/usr/lib/flink/examples/streaming/
WordCount.jar",
    "--input", "s3://myBucket/pg11.txt", "--output", "s3://myBucket/alice2/");

StepConfig flinkRunWordCount = new StepConfig()
    .withName("Flink add a wordcount step")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(flinkWordCountConf);

stepConfigs.add(flinkRunWordCount);

AddJobFlowStepsResult res = emr.addJobFlowSteps(new AddJobFlowStepsRequest()
    .withJobFlowId("myClusterId"))
```



```
.withSteps(stepConfigs));
```

Exemple AWS CLI

```
aws emr add-steps --cluster-id <j-XXXXXXXX> \  
--steps Type=CUSTOM_JAR,Name=Flink_Submit_To_Long_Running,Jar=command-runner.jar,\  
Args="flink","run","-m","yarn-cluster","-yid","application_1473169569237_0002",\  
"/usr/lib/flink/examples/streaming/WordCount.jar",\  
"--input","s3://myBucket/pg11.txt","--output","s3://myBucket/alice2/" \  
--region <region-code>
```

Soumission d'une tâche Flink transitoire

Les exemples suivants lancent un cluster transitoire qui exécute une tâche Flink, puis la résilie à son terme.

Exemple SDK pour Java

```
import java.util.ArrayList;  
import java.util.List;  
import com.amazonaws.AmazonClientException;  
import com.amazonaws.auth.AWSCredentials;  
import com.amazonaws.auth.AWSStaticCredentialsProvider;  
import com.amazonaws.auth.profile.ProfileCredentialsProvider;  
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;  
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;  
import com.amazonaws.services.elasticmapreduce.model.*;  
  
public class Main_test {  
  
    public static void main(String[] args) {  
        AWSCredentials credentials_profile = null;  
        try {  
            credentials_profile = new ProfileCredentialsProvider("default").getCredentials();  
        } catch (Exception e) {  
            throw new AmazonClientException(  
                "Cannot load credentials from .aws/credentials file. " +  
                "Make sure that the credentials file exists and the profile name is specified  
within it.",  
                e);  
        }  
  
        AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
```

```
.withCredentials(new AWSStaticCredentialsProvider(credentials_profile))
.withRegion(Regions.US_WEST_1)
.build();

List<StepConfig> stepConfigs = new ArrayList<StepConfig>();
HadoopJarStepConfig flinkWordCountConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("bash", "-c", "flink", "run", "-m", "yarn-cluster", "-yn", "2",
        "/usr/lib/flink/examples/streaming/WordCount.jar", "--input", "s3://path/to/
input-file.txt", "--output",
        "s3://path/to/output/");

StepConfig flinkRunWordCountStep = new StepConfig()
    .withName("Flink add a wordcount step and terminate")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(flinkWordCountConf);

stepConfigs.add(flinkRunWordCountStep);

Application flink = new Application().withName("Flink");

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("flink-transient")
    .withReleaseLabel("emr-5.20.0")
    .withApplications(flink)
    .withServiceRole("EMR_DefaultRole")
    .withJobFlowRole("EMR_EC2_DefaultRole")
    .withLogUri("s3://path/to/my/logfiles")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2KeyName("myEc2Key")
        .withEc2SubnetId("subnet-12ab3c45")
        .withInstanceCount(3)
        .withKeepJobFlowAliveWhenNoSteps(false)
        .withMasterInstanceType("m4.large")
        .withSlaveInstanceType("m4.large"))
    .withSteps(stepConfigs);

RunJobFlowResult result = emr.runJobFlow(request);
System.out.println("The cluster ID is " + result.toString());

}

}
```

Exemple AWS CLI

Utilisez la sous-commande `create-cluster` pour créer un cluster transitoire qui se termine quand la tâche Flink est terminée :

```
aws emr create-cluster --release-label emr-5.2.1 \  
--name "Flink_Transient" \  
--applications Name=Flink \  
--configurations file://./configurations.json \  
--region us-east-1 \  
--log-uri s3://myLogUri \  
--auto-terminate \  
--instance-type m5.xlarge \  
--instance-count 2 \  
--service-role EMR_DefaultRole_V2 \  
--ec2-attributes KeyName=<YourKeyName>,InstanceProfile=EMR_EC2_DefaultRole \  
--steps Type=CUSTOM_JAR,Jar=command-runner.jar,Name=Flink_Long_Running_Session,\  
Args="bash","-c","\"flink run -m yarn-cluster /usr/lib/flink/examples/streaming/  
WordCount.jar  
--input s3://myBucket/pg11.txt --output s3://myBucket/alice/""
```

Utilisation du shell Scala

Le shell Scala Flink pour les clusters EMR n'est configuré que pour démarrer de nouvelles sessions de YARN. Vous pouvez utiliser le shell Scala en suivant la procédure ci-dessous.

Utilisation du shell Flink Scala sur le nœud primaire

1. Connectez-vous au nœud primaire avec SSH comme décrit dans [Connexion au nœud primaire avec SSH](#).
2. Tapez ce qui suit pour démarrer un shell :

Dans Amazon EMR version 5.5.0 et versions ultérieures, vous pouvez utiliser la commande suivante pour démarrer un cluster Yarn pour le Scala Shell avec un seul TaskManager.

```
% flink-scala-shell yarn 1
```

Dans les versions antérieures d'Amazon EMR, utilisez :

```
% /usr/lib/flink/bin/start-scala-shell.sh yarn 1
```

Cela démarre le shell Scala Flink pour que vous puissiez utiliser Flink de façon interactive. Tout comme avec les autres interfaces et options, vous pouvez faire évoluer la valeur de l'option `-n` utilisée dans l'exemple basé sur le nombre de tâches à exécuter à partir du shell.

Pour plus d'informations, consultez [Scala REPL](#) dans la documentation officielle d'Apache Flink.

Recherche de l'interface web Flink

Le maître d'application qui appartient à l'application Flink héberge l'interface Web de Flink. Il s'agit d'une autre méthode pour soumettre un fichier JAR en tant que tâche ou pour afficher le statut actuel d'autres tâches. L'interface Web Flink est active tant qu'une session Flink est en cours d'exécution. Si une tâche YARN de longue durée est déjà active, vous pouvez suivre les instructions de la rubrique [Connexion au nœud primaire avec SSH](#) du Guide de gestion Amazon EMR pour vous connecter à YARN ResourceManager. Par exemple, si vous avez mis en place un tunnel SSH et que vous avez activé un proxy dans votre navigateur, vous choisissez la connexion ResourceManager sous Connexions dans votre page des détails de cluster EMR.

Cluster: Development Cluster **Waiting** Cluster ready after last step completed.

Connections:



[Resource Manager ... \(View All\)](#)

Après avoir trouvé ResourceManager, sélectionnez l'application YARN qui héberge une session Flink. Cliquez sur le lien sous la colonne Tracking UI (UI de suivi).

All Applications

Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealed Nodes
2	2 GB	11.25 GB	0 B	2	8	0	1	0	0	0	0

Scheduling Resource Type	Minimum Allocation	Maximum Allocation
MEMORY]	<memory:32, vCores:1>	<memory:11520, vCores:8>

Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
Flink session with 1 TaskManagers (detached)	Apache Flink	default	Mon Oct 10 14:42:47 -0700 2016	N/A	RUNNING	UNDEFINED		ApplicationMaster

First Previous

Dans l'interface Web Flink, vous pouvez afficher la configuration, soumettre votre propre JAR personnalisé comme une tâche, ou contrôler les travaux en cours.

Apache Flink Dashboard Overview Version: 1.1.1

- Task Managers: 1
- Task Slots: 1
- Available Task Slots: 1

Total Jobs

Status	Count
Running	0
Finished	0
Canceled	0
Failed	0

Running Jobs

Start Time	End Time	Duration	Job Name	Job ID	Tasks	Status

Completed Jobs

Start Time	End Time	Duration	Job Name	Job ID	Tasks	Status

Outil de mise à l'échelle automatique Flink

Présentation

La version 6.15.0 et les versions ultérieures d'Amazon EMR prennent en charge l'outil de mise à l'échelle automatique Flink. La fonctionnalité de mise à l'échelle automatique des tâches collecte les

métriques issues de l'exécution des tâches de streaming Flink et met à l'échelle automatiquement les vertex de chaque tâche. Cela réduit la contre-pression et répond à l'objectif d'utilisation que vous avez défini.

Pour plus d'informations, voir la rubrique [Outil de mise à l'échelle automatique](#) de la documentation de l'opérateur Apache Flink Kubernetes.

Considérations

- La version 6.15.0 et les versions ultérieures d'Amazon EMR prennent en charge l'outil de mise à l'échelle automatique Flink.
- L'outil de mise à l'échelle automatique Flink est pris en charge uniquement pour les tâches de streaming.
- Seul le planificateur adaptatif est pris en charge. Le planificateur par défaut n'est pas pris en charge.
- Nous recommandons d'activer la mise à l'échelle de cluster pour permettre une mise à disposition de ressources dynamique. La mise à l'échelle gérée par Amazon EMR est préférable, car l'évaluation des métriques est effectuée toutes les 5 à 10 secondes. À cet intervalle, votre cluster peut s'adapter plus facilement à l'évolution des ressources de cluster requises.

Activation de l'outil de mise à l'échelle automatique

Pour activer l'outil de mise à l'échelle automatique Flink lorsque vous créez un cluster Amazon EMR sur EC2, procédez comme suit.

1. Créez un cluster EMR depuis la console Amazon EMR :
 - a. Choisissez Amazon EMR version `emr-6.15.0` ou ultérieure. Sélectionnez l'offre d'applications Flink, puis toutes les autres applications que vous souhaitez inclure dans votre cluster.

Name and applications [Info](#)








Name

Amazon EMR release [Info](#)

A release contains a set of applications which can be installed on your cluster.

emr-6.15.0 ▼

Application bundle

<p>Spark Interactive</p> 	<p>Core Hadoop</p> 	<p>Flink</p> 	<p>HBase</p> 	<p>Presto</p> 	<p>Trino</p> 	<p>Custom</p> 
--	--	--	--	--	--	---

<input checked="" type="checkbox"/> Flink 1.17.1	<input type="checkbox"/> Ganglia 3.7.2	<input type="checkbox"/> HBase 2.4.17
<input type="checkbox"/> HCatalog 3.1.3	<input checked="" type="checkbox"/> Hadoop 3.3.6	<input checked="" type="checkbox"/> Hive 3.1.3
<input type="checkbox"/> Hue 4.11.0	<input type="checkbox"/> JupyterEnterpriseGateway 2.6.0	<input type="checkbox"/> JupyterHub 1.5.0
<input type="checkbox"/> Livy 0.7.1	<input type="checkbox"/> MXNet 1.9.1	<input type="checkbox"/> Oozie 5.2.1
<input type="checkbox"/> Phoenix 5.1.3	<input type="checkbox"/> Pig 0.17.0	<input type="checkbox"/> Presto 0.283
<input type="checkbox"/> Spark 3.4.1	<input type="checkbox"/> Sqoop 1.4.7	<input type="checkbox"/> TensorFlow 2.11.0
<input type="checkbox"/> Tez 0.10.2	<input type="checkbox"/> Trino 426	<input checked="" type="checkbox"/> Zeppelin 0.10.1
<input type="checkbox"/> ZooKeeper 3.5.10		

- b. Dans Mise à l'échelle et provisionnement du cluster, sélectionnez l'option Utiliser la mise à l'échelle gérée par EMR.

Cluster scaling and provisioning [Info](#)

Set up scaling and provisioning configurations for the core and task node groups for your cluster.

Choose an option

<input type="radio"/> Set cluster size manually Use this option if you know your workload patterns in advance.	<input checked="" type="radio"/> Use EMR-managed scaling Monitor key workload metrics so that EMR can optimize the cluster size and resource utilization.	<input type="radio"/> Use custom automatic scaling To programmatically scale core and task nodes, create custom automatic scaling policies.
--	---	---

2. Dans la section Paramètres du logiciel, entrez la configuration suivante pour activer l'outil de mise à l'échelle automatique Flink. Pour les scénarios de test, définissez l'intervalle de décision, l'intervalle entre les fenêtres de mesures et l'intervalle de stabilisation sur une valeur inférieure afin que la tâche prenne immédiatement une décision de mise à l'échelle et faciliter ainsi la vérification.

[

```
{
  "Classification": "flink-conf",
  "Properties": {
    "job.autoscaler.enabled": "true",
    "jobmanager.scheduler": "adaptive",
    "job.autoscaler.stabilization.interval": "60s",
    "job.autoscaler.metrics.window": "60s",
    "job.autoscaler.decision.interval": "10s",
    "job.autoscaler.debug.logs.interval": "60s"
  }
}
```

3. Sélectionnez ou configurez les autres paramètres selon vos préférences, puis créez le cluster compatible avec l'outil de mise à l'échelle automatique Flink.

Configurations de l'outil de mise à l'échelle automatique

Cette section couvre la plupart des configurations que vous pouvez modifier en fonction de vos besoins spécifiques.

Note

Avec des configurations temporelles telles que les paramètres `time`, `interval` et `window`, l'unité par défaut lorsqu'aucune unité n'est spécifiée est la milliseconde. Une valeur de 30 sans suffixe est donc égale à 30 millisecondes. Pour les autres unités de temps, incluez le suffixe approprié de `s` pour les secondes, `m` pour les minutes ou `h` pour les heures.

Rubriques

- [Configurations de boucle de l'outil de mise à l'échelle automatique](#)
- [Agrégation des métriques et configurations de l'historique](#)
- [Configurations au niveau des vertex de tâche](#)
- [Configurations de traitement du backlog](#)
- [Configurations des opérations de mise à l'échelle](#)

Configurations de boucle de l'outil de mise à l'échelle automatique

L'outil de mise à l'échelle automatique récupère les métriques au niveau des vertex de la tâche pour chaque intervalle de temps configurable, les convertit en variables d'échelle, estime le parallélisme des nouveaux vertex de la tâche et le recommande au planificateur de tâches. Les métriques ne sont collectées qu'à l'issue du redémarrage de la tâche et de l'intervalle de stabilisation du cluster.

Clé de configuration	Valeur par défaut	Description	Exemples de valeur
<code>job.autoscaler.enabled</code>	<code>false</code>	Activez la mise à l'échelle automatique sur votre cluster Flink.	<code>true</code> , <code>false</code>
<code>job.autoscaler.decision.interval</code>	<code>60s</code>	Intervalle de décision de l'outil de mise à l'échelle automatique.	<code>30</code> (l'unité par défaut est la milliseconde), <code>5m</code> , <code>1h</code>
<code>job.autoscaler.restart.time</code>	<code>3m</code>	Temps de redémarrage prévu à utiliser jusqu'à ce que l'opérateur puisse le déterminer de manière fiable à partir de l'historique.	<code>30</code> (l'unité par défaut est la milliseconde), <code>5m</code> , <code>1h</code>
<code>job.autoscaler.stabilization.interval</code>	<code>300s</code>	Période de stabilisation au cours de laquelle aucune nouvelle mise à l'échelle ne sera exécutée.	<code>30</code> (l'unité par défaut est la milliseconde), <code>5m</code> , <code>1h</code>
<code>job.autoscaler.debug.logs.interval</code>	<code>300s</code>	Intervalle entre les journaux de débogage de l'outil de mise à l'échelle automatique.	<code>30</code> (l'unité par défaut est la milliseconde), <code>5m</code> , <code>1h</code>

Agrégation des métriques et configurations de l'historique

L'outil de mise à l'échelle automatique récupère les métriques et les agrège sur une fenêtre temporelle coulissante. Les métriques sont alors évaluées dans le cadre de décisions de mise à l'échelle. L'historique des décisions de mise à l'échelle pour chaque vertex de tâche est utilisé pour estimer le nouveau parallélisme. Les métriques présentent une date d'expiration basée sur l'ancienneté et une taille d'historique (1 au minimum).

Clé de configuration	Valeur par défaut	Description	Exemples de valeur
<code>job.autoscaler.metrics.window</code>	600s	Scaling metrics aggregation window size.	30 (l'unité par défaut est la milliseconde), 5m, 1h
<code>job.autoscaler.history.max.count</code>	3	Nombre maximal de décisions de mise à l'échelle passées à conserver par vertex.	1 sur Integer.MAX_VALUE
<code>job.autoscaler.history.max.age</code>	24h	Nombre minimal de décisions de mise à l'échelle passées à conserver par vertex.	30 (l'unité par défaut est la milliseconde), 5m, 1h

Configurations au niveau des vertex de tâche

Le parallélisme de chaque vertex de tâche est modifié en fonction de l'utilisation de la cible et est limité par les limites minimale et maximale de parallélisme. Nous vous déconseillons de définir une cible d'utilisation proche de 100 % (c'est-à-dire une valeur de 1). La limite d'utilisation joue le rôle de tampon pour gérer les fluctuations de charge intermédiaires.

Clé de configuration	Valeur par défaut	Description	Exemples de valeur
<code>job.autoscaler.target.utilization</code>	0.7	Utilisation des vertex cibles.	0 - 1

Clé de configuration	Valeur par défaut	Description	Exemples de valeur
<code>job.autoscaler.target.utilization.boundary</code>	0.4	Limite d'utilisation des vertex cibles. La mise à l'échelle ne sera pas effectuée si le taux de traitement actuel est compris entre $[\text{target_rate} / (\text{target_utilization} - \text{boundary})]$ et $[\text{target_rate} / (\text{target_utilization} + \text{boundary})]$.	0 - 1
<code>job.autoscaler.vertex.min-parallelism</code>	1	Parallélisme minimal que l'outil de mise à l'échelle automatique peut utiliser.	0 - 200
<code>job.autoscaler.vertex.max-parallelism</code>	200	Parallélisme maximal que l'outil de mise à l'échelle automatique peut utiliser. Notez que l'outil de mise à l'échelle automatique ignore cette limite si elle est supérieure au parallélisme maximal configuré dans la configuration Flink ou directement sur chaque opérateur.	0 - 200

Configurations de traitement du backlog

Le vertex de la tâche requiert des ressources supplémentaires pour gérer les événements en attente, ou backlogs, qui s'accumulent pendant le délai de mise à l'échelle. C'est ce que l'on appelle également la durée de `catch-up`. Si le délai de traitement du backlog dépasse la `lag - threshold` valeur configurée, l'utilisation cible du vertex de la tâche augmente jusqu'au niveau maximum. Cela permet d'éviter les opérations de mise à l'échelle inutiles pendant le traitement du backlog.

Clé de configuration	Valeur par défaut	Description	Exemples de valeur
<code>job.autoscaler.backlog-processing.lag-threshold</code>	5m	Seuil de retard qui évitera les mises à l'échelle inutiles tout en supprimant les messages en attente responsables du retard.	30 (l'unité par défaut est la milliseconde), 5m, 1h
<code>job.autoscaler.catch-up.duration</code>	15m	Durée cible pour le traitement complet de tout backlog après une opération de mise à l'échelle. Définissez cette valeur sur 0 pour désactiver la mise à l'échelle basée sur le backlog.	30 (l'unité par défaut est la milliseconde), 5m, 1h

Configurations des opérations de mise à l'échelle

Aucune opération de réduction n'est effectuée par l'outil de mise à l'échelle automatique immédiatement après une opération de mise à l'échelle pendant le délai de grâce. Cela permet d'éviter tout cycle inutile d'opérations successives d'augmentation ou de réduction d'échelle provoquées par des fluctuations de charge temporaires.

Nous pouvons utiliser le ratio de réduction d'échelle pour réduire progressivement le parallélisme et libérer des ressources pour faire face à des pics de charge temporaires. Cela permet également d'éviter des opérations d'augmentation d'échelle mineures et inutiles après une opération de réduction majeure.

Nous pouvons détecter une opération de mise à l'échelle inefficace sur la base de l'historique des décisions d'augmentation d'échelle des vertex de tâches passées afin d'empêcher toute nouvelle modification du parallélisme.

Clé de configuration	Valeur par défaut	Description	Exemples de valeur
<code>job.autoscaler.scale-up.grace-period</code>	1h	Durée pendant laquelle aucune réduction d'échelle d'un vertex n'est autorisée après l'augmentation d'échelle de ce vertex.	30 (l'unité par défaut est la milliseconde), 5m, 1h
<code>job.autoscaler.scale-down.max-factor</code>	0.6	Facteur maximal de réduction d'échelle. Une valeur égale à 1 signifie qu'aucune limite de réduction d'échelle n'est définie ; 0.6 indique que seule une réduction d'échelle de 60 % du parallélisme d'origine peut être appliquée à la tâche.	0 - 1
<code>job.autoscaler.scale-up.max-factor</code>	100000.	Ratio maximal d'augmentation d'échelle. Une valeur de 2.0 signifie que seule une augmentation d'échelle de 200 % du parallélisme actuel peut être appliquée à la tâche.	0 - Integer.MAX_VALUE
<code>job.autoscaler.scaling.effectiveness.detection.enabled</code>	false	Permet d'activer la détection des opérations de mise à l'échelle inefficaces et d'autoriser le blocage de nouvelles mises à l'échelle par l'outil de mise à l'échelle automatique.	true, false

Optimisation des temps de redémarrage pour les opérations de récupération et de mise à l'échelle de tâches

Lorsqu'une tâche échoue ou qu'une opération de mise à l'échelle est en cours, Flink tente de réexécuter la tâche à partir du dernier point de contrôle terminé. L'exécution du processus de redémarrage peut durer une minute ou plus, en fonction de la taille de l'état du point de contrôle et du nombre de tâches parallèles. Pendant la période de redémarrage, les tâches en attente peuvent s'accumuler pour la tâche. Flink peut cependant permettre d'optimiser la vitesse de récupération et de redémarrage des graphes d'exécution afin d'améliorer la stabilité des tâches.

Cette page décrit comment Amazon EMR Flink peut améliorer le temps de redémarrage des tâches lors des opérations de récupération ou de mise à l'échelle de tâches.

Rubriques

- [Récupération locale des tâches](#)
- [Point de contrôle incrémentiel générique basé sur les journaux](#)
- [Récupération précise](#)
- [Mécanisme de redémarrage combiné dans le planificateur adaptatif](#)

Récupération locale des tâches

Note

La récupération locale des tâches est prise en charge avec Amazon EMR 6.0.0 et versions ultérieures.

Avec les points de contrôle Flink, chaque tâche crée un instantané de son état, que Flink écrit sur un système de stockage distribué tel qu'Amazon S3. En cas de récupération, les tâches restaurent leur état à partir du stockage distribué. Le stockage distribué offre une tolérance aux pannes et peut redistribuer l'état lors de la mise à l'échelle, car tous les nœuds peuvent y accéder.

Cependant, un magasin distribué à distance présente également un inconvénient : toutes les tâches doivent lire leur état depuis un emplacement distant sur le réseau, ce qui peut entraîner l'augmentation du temps de récupération pour les états importants lors des opérations de récupération ou de mise à l'échelle de tâches.

La récupération locale des tâches permet de résoudre ce problème. Les tâches enregistrent leur état au point de contrôle sur un stockage secondaire local à la tâche, par exemple sur un disque local. Elles stockent également leur état sur le stockage principal, à savoir Amazon S3 dans notre cas. Pendant la récupération, le planificateur planifie les tâches sur le même Task Manager que celui dans lequel les tâches ont été exécutées précédemment afin qu'elles puissent être récupérées depuis le magasin d'état local au lieu de lire depuis le magasin d'état distant. Pour plus d'informations, voir la rubrique [Récupération locale des tâches](#) de la documentation Apache Flink.

Nos tests d'évaluation avec des exemples de tâches ont montré que le temps de récupération était passé de quelques minutes à quelques secondes grâce à l'activation de la récupération locale des tâches.

Pour activer la récupération locale des tâches, définissez les configurations suivantes dans votre fichier `flink-conf.yaml`. Spécifiez la valeur de l'intervalle de point de contrôle en millisecondes.

```
state.backend.local-recovery: true
state.backend: hasmap or rocksdb
state.checkpoints.dir: s3://storage-location-bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

Point de contrôle incrémentiel générique basé sur les journaux

Note

Le point de contrôle incrémentiel générique basé sur les journaux est pris en charge avec Amazon EMR 6.10.0 et versions ultérieures.

Le point de contrôle incrémentiel générique basé sur les journaux a été ajouté dans Flink 1.16 pour accélérer les points de contrôle. Un intervalle de point de contrôle plus court entraîne souvent une réduction du travail de récupération, car moins d'événements doivent être traités de nouveau après la récupération. Pour plus d'informations, accédez à la page [Improving speed and stability of checkpointing with generic log-based incremental checkpoints](#) sur le blog Apache Flink.

Sur base de quelques exemples de tâches, nos tests d'évaluation ont montré que le temps de contrôle était passé de quelques minutes à quelques secondes grâce au point de contrôle incrémentiel générique basé sur les journaux.

Pour activer les points de contrôle incrémentiels génériques basés sur les journaux, définissez les configurations suivantes dans votre fichier `flink-conf.yaml`. Spécifiez la valeur de l'intervalle de point de contrôle en millisecondes.

```
state.backend.changelog.enabled: true
state.backend.changelog.storage: filesystem
dstl.dfs.base-path: s3://bucket-path/changelog
state.backend.local-recovery: true
state.backend: rocksdb
state.checkpoints.dir: s3://bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

Récupération précise

Note

La prise en charge de la récupération précise pour le planificateur par défaut est disponible avec Amazon EMR 6.0.0 et versions ultérieures. La prise en charge de la récupération précise pour le planificateur adaptatif est disponible avec Amazon EMR 6.15.0 et versions ultérieures.

Lorsqu'une tâche échoue pendant son exécution, Flink réinitialise l'intégralité du graphe d'exécution et déclenche une réexécution complète à partir du dernier point de contrôle terminé. Cette opération est plus chère qu'une simple réexécution des tâches qui ont échoué. La récupération précise redémarre uniquement le composant connecté au pipeline de la tâche ayant échoué. Dans l'exemple suivant, le graphe de tâches présente 5 sommets (A à E). Toutes les connexions entre les sommets sont en pipeline avec une distribution ponctuelle, et la valeur de `parallelism.default` pour la tâche est définie sur 2.

```
A # B # C # D # E
```

Dans cet exemple, 10 tâches sont en cours d'exécution au total. Le premier pipeline (a1 à e1) s'exécute sur un TaskManager (TM1), et le deuxième pipeline (a2 à e2) s'exécute sur un autre TaskManager (TM2).

```
a1 # b1 # c1 # d1 # e1
a2 # b2 # c2 # d2 # e2
```


Deux composants sont connectés en pipeline : a1 # e1 et a2 # e2. En cas d'échec de TM1 ou de TM2, l'échec affecte uniquement les 5 tâches du pipeline dans lequel TaskManager était en cours d'exécution. La stratégie de redémarrage démarre uniquement le composant en pipeline concerné.

La récupération précise ne fonctionne qu'avec des tâches Flink parfaitement parallèles. Elle n'est pas prise en charge par `keyBy()` ou par les opérations `redistribute()`. Pour plus d'informations, accédez à la page [FLIP-1 : Fine Grained Recovery from Task Failures](#) du projet Jira Flink Improvement Proposal.

Pour activer la récupération précise, définissez les configurations suivantes dans votre fichier `flink-conf.yaml`.

```
jobmanager.execution.failover-strategy: region
restart-strategy: exponential-delay or fixed-delay
```

Mécanisme de redémarrage combiné dans le planificateur adaptatif

Note

Le mécanisme de redémarrage combiné du planificateur adaptatif est pris en charge par Amazon EMR 6.15.0 et versions ultérieures.

Le planificateur adaptatif peut ajuster le parallélisme de la tâche en fonction des emplacements disponibles. Si le nombre d'emplacements disponibles est insuffisant, le planificateur réduit automatiquement le parallélisme pour s'adapter au parallélisme des tâches configuré. Si de nouveaux emplacements sont disponibles, la tâche fait l'objet d'une augmentation d'échelle selon le parallélisme des tâches configuré. Un planificateur adaptatif permet d'éviter les temps d'arrêt de la tâche lorsque les ressources disponibles sont insuffisantes. Il s'agit du planificateur pris en charge par l'outil de mise à l'échelle automatique Flink. Nous recommandons donc l'utilisation d'un planificateur adaptatif avec Amazon EMR Flink. Les planificateurs adaptatifs peuvent toutefois effectuer plusieurs redémarrages en peu de temps, à raison d'un redémarrage pour chaque nouvelle ressource ajoutée, ce qui peut entraîner une baisse des performances de la tâche.

Avec Amazon EMR 6.15.0 et versions ultérieures, Flink dispose d'un mécanisme de redémarrage combiné dans le planificateur adaptatif qui ouvre une fenêtre de redémarrage lorsque la première ressource est ajoutée, puis attend l'intervalle de fenêtre configuré de 1 minute par défaut. Un seul redémarrage est effectué lorsque les ressources disponibles sont suffisantes pour exécuter la tâche avec un parallélisme configuré ou lorsque l'intervalle expire.

Grâce à quelques exemples de tâches, nos tests d'évaluation ont montré que cette fonctionnalité traite 10 % d'enregistrements supplémentaires par rapport au comportement par défaut lorsque vous utilisez le planificateur adaptatif et l'outil de mise à l'échelle automatique Flink.

Pour activer le mécanisme de redémarrage combiné, définissez les configurations suivantes dans votre fichier `flink-conf.yaml`.

```
jobmanager.adaptive-scheduler.combined-restart.enabled: true
jobmanager.adaptive-scheduler.combined-restart.window-interval: 1m
```

Travailler avec les jobs Flink de Zeppelin dans Amazon EMR

Introduction

Les versions 6.10.0 et supérieures d'Amazon EMR prennent en charge l'intégration [Apache Zeppelin](#) avec Apache Flink. Vous pouvez soumettre des jobs Flink de manière interactive via les blocs-notes Zeppelin. Avec l'interpréteur Flink, vous pouvez exécuter des requêtes Flink, définir des tâches de streaming et de traitement par lots Flink, et visualiser le résultat dans les blocs-notes Zeppelin. L'interpréteur Flink est basé sur Flink REST API. Cela vous permet d'accéder aux tâches Flink et de les manipuler depuis l'environnement Zeppelin pour effectuer un traitement et une analyse des données en temps réel.

Il existe quatre sous-interprètes dans Flink Interpreter. Ils ont des objectifs différents, mais ils se trouvent tous dans la JVM et partagent les mêmes points d'entrée préconfigurés vers Flink (`ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment`). Les interprètes sont les suivants :

- `%flink` – Crée `ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment`, et fournit un environnement Scala
- `%flink.pyflink` – Fournit un environnement Python
- `%flink.ssql` – Fournit un environnement SQL de streaming
- `%flink.bsqli` – Fournit un environnement SQL par lots

Prérequis

- L'intégration de Zeppelin à Flink est prise en charge pour les clusters créés avec Amazon EMR 6.10.0 et versions ultérieures.

- Pour afficher les interfaces Web hébergées sur des clusters EMR conformément à ces étapes, vous devez configurer un tunnel SSH pour autoriser l'accès entrant. Pour plus d'informations, consultez la rubrique [Configuration des paramètres de proxy pour afficher les sites web hébergés sur le nœud primaire](#).

Configurer Zeppelin-Flink sur un cluster EMR

Procédez comme suit pour configurer Apache Flink sur Apache Zeppelin afin qu'il s'exécute sur un cluster EMR :

1. Créez un nouveau cluster depuis la console Amazon EMR. Sélectionnez emr-6.10.0 ou supérieur pour la version Amazon EMR. Choisissez ensuite de personnaliser votre bundle d'applications avec l'option Personnaliser. Incluez au moins Flink, Hadoop et Zeppelin dans votre offre groupée.

Amazon EMR release [Info](#)
 A release contains a set of applications which can be installed on your cluster.

emr-6.10.0 ▼

Application bundle

Spark

Core Hadoop

HBase

Presto

Trino

Custom

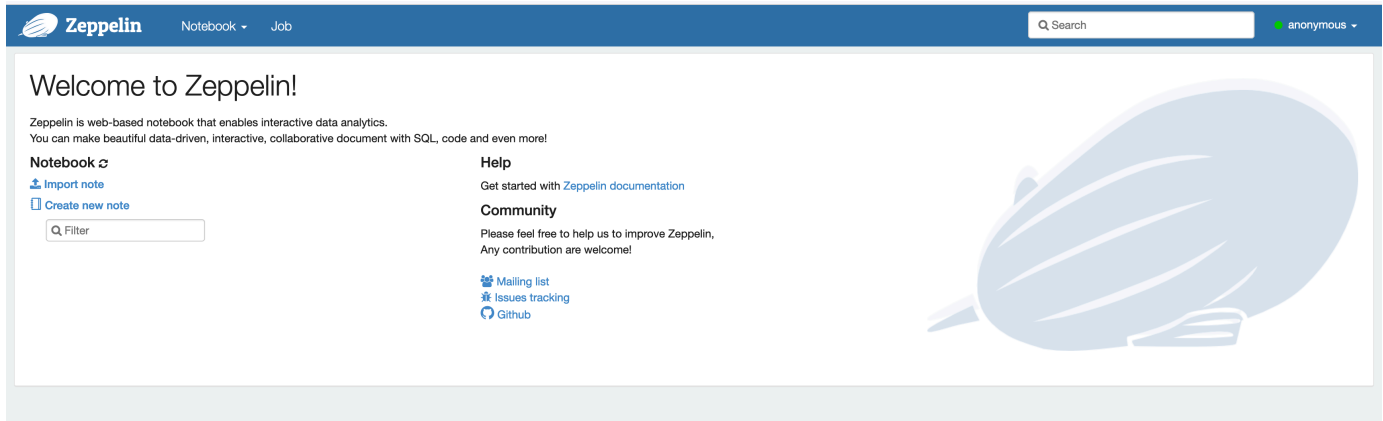
▼ Customize your application bundle

Applications included in bundle

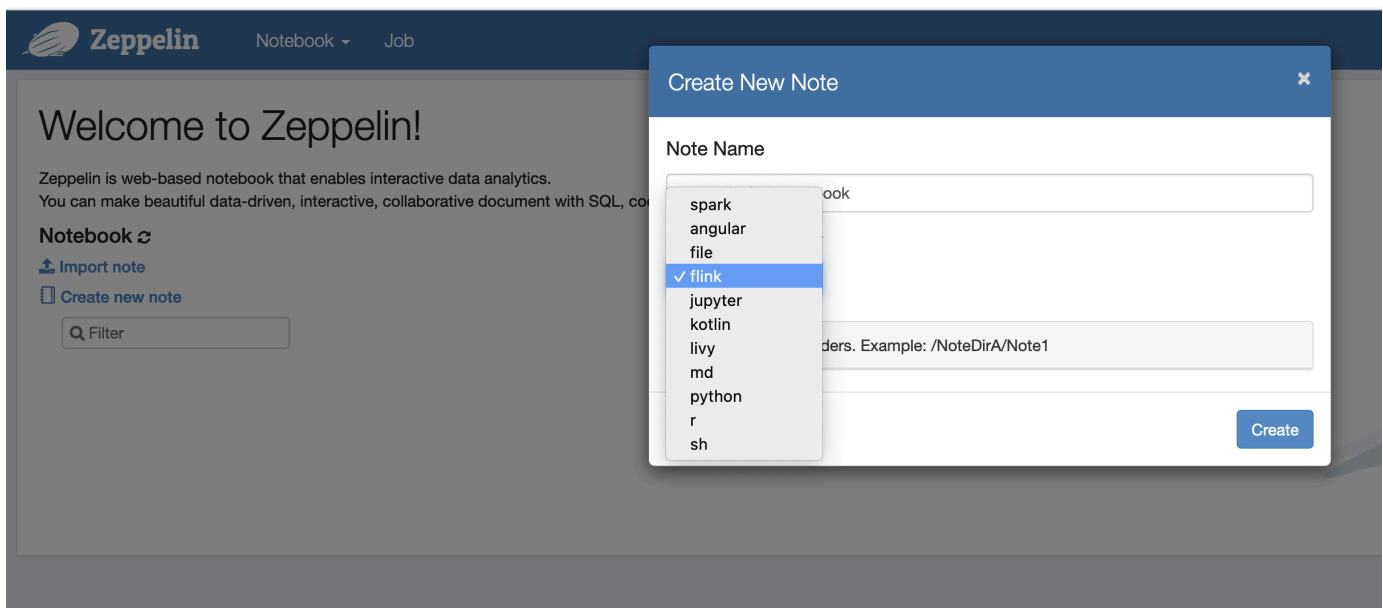
<input checked="" type="checkbox"/> Flink 1.16.0	<input type="checkbox"/> Ganglia 3.7.2
<input type="checkbox"/> HBase 2.4.15	<input type="checkbox"/> HCatalog 3.1.3
<input checked="" type="checkbox"/> Hadoop 3.3.3	<input type="checkbox"/> Hive 3.1.3
<input type="checkbox"/> Hue 4.10.0	<input type="checkbox"/> JupyterEnterpriseGateway 2.6.0
<input type="checkbox"/> JupyterHub 1.5.0	<input type="checkbox"/> Livy 0.7.1
<input type="checkbox"/> MXNet 1.9.1	<input type="checkbox"/> Oozie 5.2.1
<input type="checkbox"/> Phoenix 5.1.2	<input type="checkbox"/> Pig 0.17.0
<input type="checkbox"/> Presto 0.278	<input type="checkbox"/> Spark 3.3.1
<input type="checkbox"/> Sqoop 1.4.7	<input type="checkbox"/> TensorFlow 2.11.0
<input type="checkbox"/> Tez 0.10.2	<input type="checkbox"/> Trino 403
<input checked="" type="checkbox"/> Zeppelin 0.10.1	<input type="checkbox"/> ZooKeeper 3.5.10

2. Créez le reste de votre cluster avec les paramètres que vous préférez.

- Une fois que votre cluster est en cours d'exécution, sélectionnez-le dans la console pour afficher ses détails et ouvrez l'onglet Applications. Sélectionnez Zeppelin dans la section Interfaces utilisateur de l'application pour ouvrir l'interface Web Zeppelin. Assurez-vous d'avoir configuré l'accès à l'interface Web Zeppelin avec un tunnel SSH vers le nœud primaire et une connexion proxy, comme décrit dans le [Prérequis](#).



- Vous pouvez désormais créer une nouvelle note dans un carnet Zeppelin avec Flink comme interpréteur par défaut.



- Reportez-vous aux exemples de code suivants qui montrent comment exécuter des tâches Flink à partir d'un bloc-notes Zeppelin.

Exécuter des tâches Flink avec Zeppelin-Flink sur un cluster EMR

- Exemple 1, Flink Scala

a) WordCount Exemple de lot (SCALA)

```
%flink

val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .groupBy(0)
    .sum(1)
    .print()
```

b) WordCount Exemple de streaming (SCALA)

```
%flink

val data = senv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .keyBy(0)
    .sum(1)
    .print

senv.execute()
```

Batch WordCount FLINK JOB FINISHED

```
%flink

val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .groupBy(0)
    .sum(1)
    .print()

data: org.apache.flink.api.scala.DataSet[String] = org.apache.flink.api.scala.DataSet@22fe7dd5
(flink,1)
(hadoop,1)
(hello,3)
(world,1)

Took 56 sec. Last updated by anonymous at May 04 2023, 2:19:24 PM. (updated)
```

Streaming WordCount FLINK JOB FINISHED

```
%flink

val data = senv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .keyBy(0)
    .sum(1)
    .print

senv.execute()

data: org.apache.flink.streaming.api.scala.DataStream[String] = org.apache.flink.streaming.api.scala.DataStream@282340f2
warning: there was one deprecation warning; for details, enable ':setting -deprecation' or ':replay -deprecation'
res2: org.apache.flink.streaming.api.datastream.DataStreamSink[(String, Int)] = org.apache.flink.streaming.api.d
atostream.DataStreamSink@7a0e7b2f
res3: org.apache.flink.api.common.JobExecutionResult =
Program execution finished
Job with JobID a78d1a9b33c0fb71b8d53f00cda51030 has finished.
Job Runtime: 8485 ms

Took 12 sec. Last updated by anonymous at May 04 2023, 2:21:05 PM. (updated)
```

- Exemple 2, Flink Streaming SQL

```
%flink.sql
SET 'sql-client.execution.result-mode' = 'tableau';
SET 'table.dml-sync' = 'true';
SET 'execution.runtime-mode' = 'streaming';
```

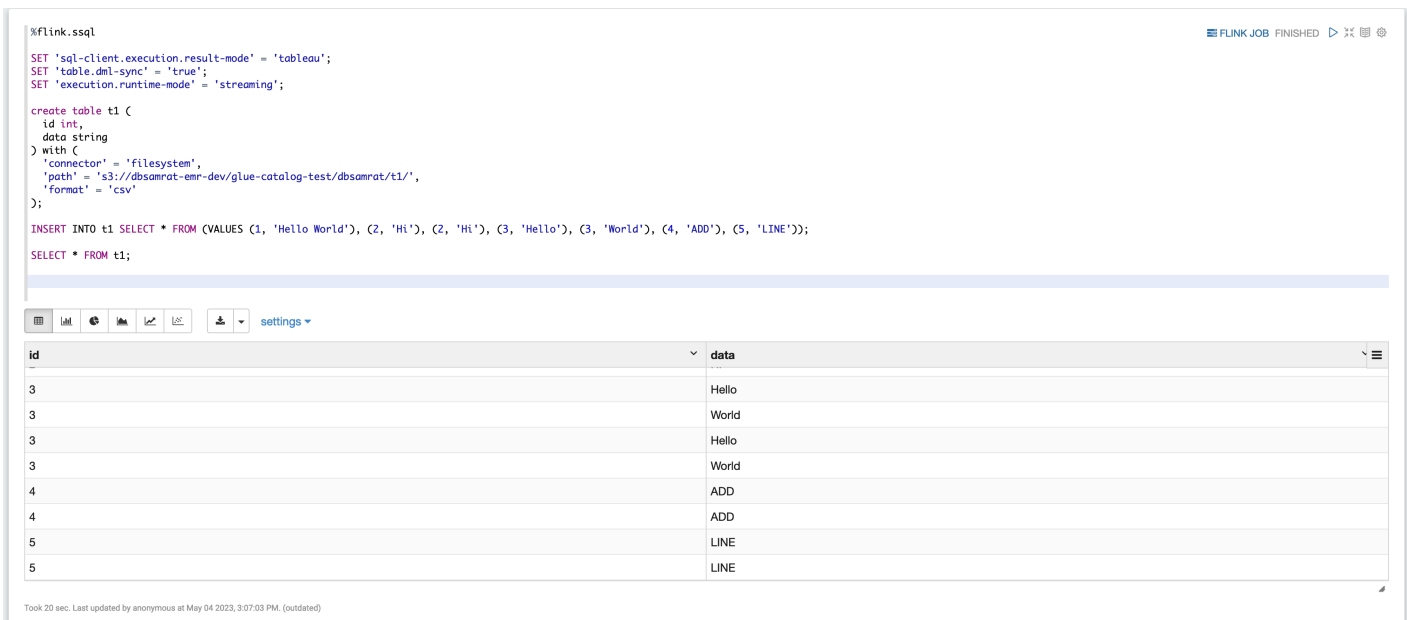
```

create table dummy_table (
  id int,
  data string
) with (
  'connector' = 'filesystem',
  'path' = 's3://s3-bucket/dummy_table',
  'format' = 'csv'
);

INSERT INTO dummy_table SELECT * FROM (VALUES (1, 'Hello World'), (2, 'Hi'), (2,
'Hi'), (3, 'Hello'), (3, 'World'), (4, 'ADD'), (5, 'LINE'));

SELECT * FROM dummy_table;

```



```

%flink.sql
SET 'sql-client.execution.result-mode' = 'tableau';
SET 'table.dml-sync' = 'true';
SET 'execution.runtime-mode' = 'streaming';

create table t1 (
  id int,
  data string
) with (
  'connector' = 'filesystem',
  'path' = 's3://dbsamrat-emr-dev/glue-catalog-test/dbsamrat/t1/',
  'format' = 'csv'
);

INSERT INTO t1 SELECT * FROM (VALUES (1, 'Hello World'), (2, 'Hi'), (2, 'Hi'), (3, 'Hello'), (3, 'World'), (4, 'ADD'), (5, 'LINE'));

SELECT * FROM t1;

```

id	data
3	Hello
3	World
3	Hello
3	World
4	ADD
4	ADD
5	LINE
5	LINE

Took 20 sec. Last updated by anonymous at May 04 2023, 3:07:03 PM. (outdated)

- Exemple 3, Pyflink. Notez que vous devez télécharger votre propre exemple de fichier texte nommé `word.txt` dans votre compartiment S3.

```

%flink.pyflink

import argparse
import logging
import sys

from pyflink.common import Row
from pyflink.table import (EnvironmentSettings, TableEnvironment, TableDescriptor,
Schema,
                             DataTypes, FormatDescriptor)

```

```
from pyflink.table.expressions import lit, col
from pyflink.table.udf import udtf

def word_count(input_path, output_path):
    t_env = TableEnvironment.create(EnvironmentSettings.in_streaming_mode())
    # write all the data to one file
    t_env.get_config().set("parallelism.default", "1")

    # define the source
    if input_path is not None:
        t_env.create_temporary_table(
            'source',
            TableDescriptor.for_connector('filesystem')
                .schema(Schema.new_builder()
                    .column('word', DataTypes.STRING())
                    .build())
                .option('path', input_path)
                .format('csv')
                .build())
        tab = t_env.from_path('source')
    else:
        print("Executing word_count example with default input data set.")
        print("Use --input to specify file input.")
        tab = t_env.from_elements(map(lambda i: (i,), word_count_data),
            DataTypes.ROW([DataTypes.FIELD('line',
DataTypes.STRING())]))

    # define the sink
    if output_path is not None:
        t_env.create_temporary_table(
            'sink',
            TableDescriptor.for_connector('filesystem')
                .schema(Schema.new_builder()
                    .column('word', DataTypes.STRING())
                    .column('count', DataTypes.BIGINT())
                    .build())
                .option('path', output_path)
                .format(FormatDescriptor.for_format('canal-json')
                    .build())
                .build())
    else:
        print("Printing result to stdout. Use --output to specify output path.")
        t_env.create_temporary_table(
            'sink',
```

```

TableDescriptor.for_connector('print')
    .schema(Schema.new_builder()
        .column('word', DataTypes.STRING())
        .column('count', DataTypes.BIGINT())
        .build())
    .build()

@udtf(result_types=[DataTypes.STRING()])
def split(line: Row):
    for s in line[0].split():
        yield Row(s)

# compute word count
tab.flat_map(split).alias('word') \
    .group_by(col('word')) \
    .select(col('word'), lit(1).count) \
    .execute_insert('sink') \
    .wait()

logging.basicConfig(stream=sys.stdout, level=logging.INFO, format="%(message)s")

word_count("s3://s3_bucket/word.txt", "s3://s3_bucket/demo_output.txt")

```

1. Choisissez FLINK JOB dans l'interface utilisateur Zeppelin pour accéder à l'interface utilisateur Web de Flink et l'afficher.



Batch WordCount FLINK JOB FINISHED ▶ ⌵ ⌶ ⚙

```

%flink
val data = benv.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .groupBy(0)
    .sum(1)
    .print()

data: org.apache.flink.api.scala.DataSet[String] = org.apache.flink.api.scala.DataSet@22fe7dd5
(flink,1)
(hadoop,1)
(hello,3)
(world,1)

Took 56 sec. Last updated by anonymous at May 04 2023, 2:19:24 PM. (outdated)

```


2. Lorsque vous choisissez FLINK JOB, vous accédez à la console Web Flink dans un autre onglet de votre navigateur.

The screenshot displays the Apache Flink Dashboard interface. On the left is a dark navigation sidebar with options: Overview, Jobs (selected), Running Jobs, Completed Jobs, Task Managers, Job Manager, and Submit New Job. The main content area shows:

- Available Task Slots:** A green '0' with 'Total Task Slots 0' and 'Task Managers 0' below it.
- Running Jobs:** A green '0' with 'Finished 2', 'Canceled 0', and 'Failed 0' below it.
- Running Job List:** A table with columns: Job Name, Start Time, Duration, End Time, Tasks, and Status. It currently shows 'No Data'.
- Completed Job List:** A table with columns: Job Name, Start Time, Duration, End Time, Tasks, and Status. It lists two jobs:

Job Name	Start Time	Duration	End Time	Tasks	Status
Flink Streaming Job	2023-05-04 14:20:56	8s	2023-05-04 14:21:04	2 / 2	FINISHED
Flink Java Job at Thu May 04 08:49:10 UTC 2023	2023-05-04 14:19:11	12s	2023-05-04 14:19:23	3 / 3	FINISHED

Historique des publications de Flink

Le tableau suivant répertorie la version de Flink incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Flink

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-7.2.0	1.18.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
		resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta
emr-5.36.2	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-7.1.0	1.18.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-7.0.0	1.18.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.15.0	1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-6.14.0	1.17.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.13.0	1.17.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-6.12.0	1.17.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.11.1	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-6.11.0	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors
emr-6.10.1	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.10.0	1.16.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-6.9.1	1.15.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.9.0	1.15.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.8.1	1.15.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-6.8.0	1.15.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.7.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-5.36.1	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.36.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-6.6.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-5.35.0	1.14.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-6.5.0	1.14.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.4.0	1.13.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi
emr-6.3.1	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-6.3.0	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-6.2.1	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-6.2.0	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-6.1.1	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-6.1.0	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.34.0	1.13.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.33.1	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-5.33.0	1.12.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-5.32.1	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.32.0	1.11.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-5.31.1	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config
emr-5.31.0	1.11.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.30.2	1.10.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.30.1	1.10.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.30.0	1.10.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
mr-5.29.0	1.9.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.28.1	1.9.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.28.0	1.9.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.27.1	1.8.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.27.0	1.8.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.26.0	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.25.0	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.24.1	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.24.0	1.8.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.23.1	1.7.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.23.0	1.7.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.22.0	1.7.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.21.2	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.21.1	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-ya rn-timeline-server, flink-client
emr-5.21.0	1.7.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-ya rn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.20.1	1.6.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.20.0	1.6.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.19.1	1.6.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.19.0	1.6.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.18.1	1.6.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.18.0	1.6.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.17.2	1.5.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.17.1	1.5.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.17.0	1.5.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.16.1	1.5.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.16.0	1.5.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.15.1	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.15.0	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.14.2	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.14.1	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.14.0	1.4.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.13.1	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.13.0	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.12.3	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.12.2	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.12.1	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.12.0	1.4.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.11.4	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.11.3	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.11.2	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.11.1	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.11.0	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.10.1	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.10.0	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.9.1	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.9.0	1.3.2	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.8.3	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.8.2	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.8.1	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.8.0	1.3.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.7.1	1.3.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.7.0	1.3.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.6.1	1.2.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client
emr-5.6.0	1.2.1	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.5.4	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.5.3	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.5.2	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.5.1	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.5.0	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.4.1	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.4.0	1.2.0	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.3.2	1.1.4	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.3.1	1.1.4	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.3.0	1.1.4	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.2.3	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.2.2	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.2.1	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.2.0	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client
emr-5.1.1	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Étiquette Amazon EMR Release	Version Flink	Composants installés avec Flink
emr-5.1.0	1.1.3	emrfs, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, flink-client

Notes de mise à jour de Flink par version

Consultez les sections suivantes pour les notes de mise à jour complètes.

Amazon EMR 7.2.0 - Notes de mise à jour de Flink

Type	Description
Amélioration	Support pour l'ajout d'étiquettes personnalisées au service Kubernetes par tâche Flink via la configuration. <code>kubernetes.service.labels</code>

Ganglia

Note

La dernière version d'Amazon EMR à inclure Ganglia était Amazon EMR 6.15.0. Pour surveiller votre cluster, les versions supérieures à 6.15.0 incluent l'agent [Amazon CloudWatch](#).

Le projet open source Ganglia est un système distribué évolutif, conçu pour surveiller les clusters et les grilles tout en réduisant au maximum l'impact sur leurs performances. Lorsque vous activez Ganglia sur votre cluster, vous pouvez générer des rapports et afficher la performance du cluster dans son ensemble, ainsi qu'inspecter la performance des instances de chaque nœud. Ganglia est également configuré pour intégrer et visualiser les métriques Hadoop et Spark. Pour plus d'informations sur le projet open source Ganglia, accédez à <http://ganglia.info/>.

Lorsque vous affichez l'interface utilisateur web de Ganglia dans un navigateur, vous voyez une vue d'ensemble des performances du cluster, avec des graphiques détaillant la charge, l'utilisation de la mémoire, l'utilisation de l'UC et le trafic réseau du cluster. Sous les statistiques du cluster, des graphiques sont fournis pour chaque serveur individuel figurant dans le cluster.

Le tableau suivant répertorie la version de Ganglia incluse dans la dernière version d'Amazon EMR série 6.x, ainsi que les composants qu'Amazon EMR installe avec Ganglia.

Pour connaître la version des composants installés avec Ganglia dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Ganglia pour emr-6.15.0

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-6.15.0	Ganglia 3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
		server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Le tableau suivant répertorie la version de Ganglia incluse dans la dernière version d'Amazon EMR série 5.x, ainsi que les composants qu'Amazon EMR installe avec Ganglia.

Pour la version des composants installés avec Ganglia dans cette version, voir la [version 5.36.2](#) Versions des composants.

Informations sur la version de Ganglia pour emr-5.36.2

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.36.2	Ganglia 3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Rubriques

- [Création d'un cluster à l'aide de Ganglia](#)
- [Affichage des métriques Ganglia](#)
- [Métriques Hadoop et Spark dans Ganglia](#)
- [Historique des versions de Ganglia](#)

Création d'un cluster à l'aide de Ganglia

Note

La dernière version d'Amazon EMR à inclure Ganglia était Amazon EMR 6.15.0. Pour surveiller votre cluster, les versions supérieures à 6.15.0 incluent l'agent [Amazon CloudWatch](#).

Pour créer un cluster avec Ganglia à l'aide de la console

1. Accédez à la nouvelle console Amazon EMR et sélectionnez **Changer** pour l'ancienne console depuis le menu latéral. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).
2. Choisissez **Créer un cluster**.
3. Dans **Software configuration (Configuration des logiciels)**, choisissez **All Applications (Toutes les applications)**, **Core Hadoop** ou **Spark**.
4. Procédez à la création du cluster avec les configurations qui conviennent.

Pour ajouter Ganglia à un cluster à l'aide du AWS CLI

Dans le AWS CLI, vous pouvez ajouter Ganglia à un cluster en utilisant `create-cluster` le `--applications` paramètre. Si vous spécifiez uniquement Ganglia à l'aide du paramètre `--applications`, Ganglia est la seule application installée.

- Tapez la commande suivante pour ajouter Ganglia lorsque vous créez un cluster et remplacez *myKey* par le nom de votre paire de clés EC2.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Spark cluster with Ganglia" --release-label  
emr-6.15.0; \  

```

```
--applications Name=Spark Name=Ganglia \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \  
--instance-count 3 --use-default-roles
```

Lorsque vous spécifiez le nombre d'instances sans utiliser le paramètre `--instance-groups`, un seul nœud maître est lancé et les instances restantes sont lancées en tant que nœuds principaux. Tous les nœuds utilisent le type d'instance spécifié dans la commande.

Note

Si vous n'avez pas encore créé le rôle de service EMR par défaut et le profil d'instance EC2, tapez `aws emr create-default-roles` pour les créer avant de taper la sous-commande `create-cluster`.

Pour plus d'informations sur l'utilisation des commandes Amazon EMR dans le AWS CLI, consultez. <https://docs.aws.amazon.com/cli/latest/reference/emr>

Affichage des métriques Ganglia

Note

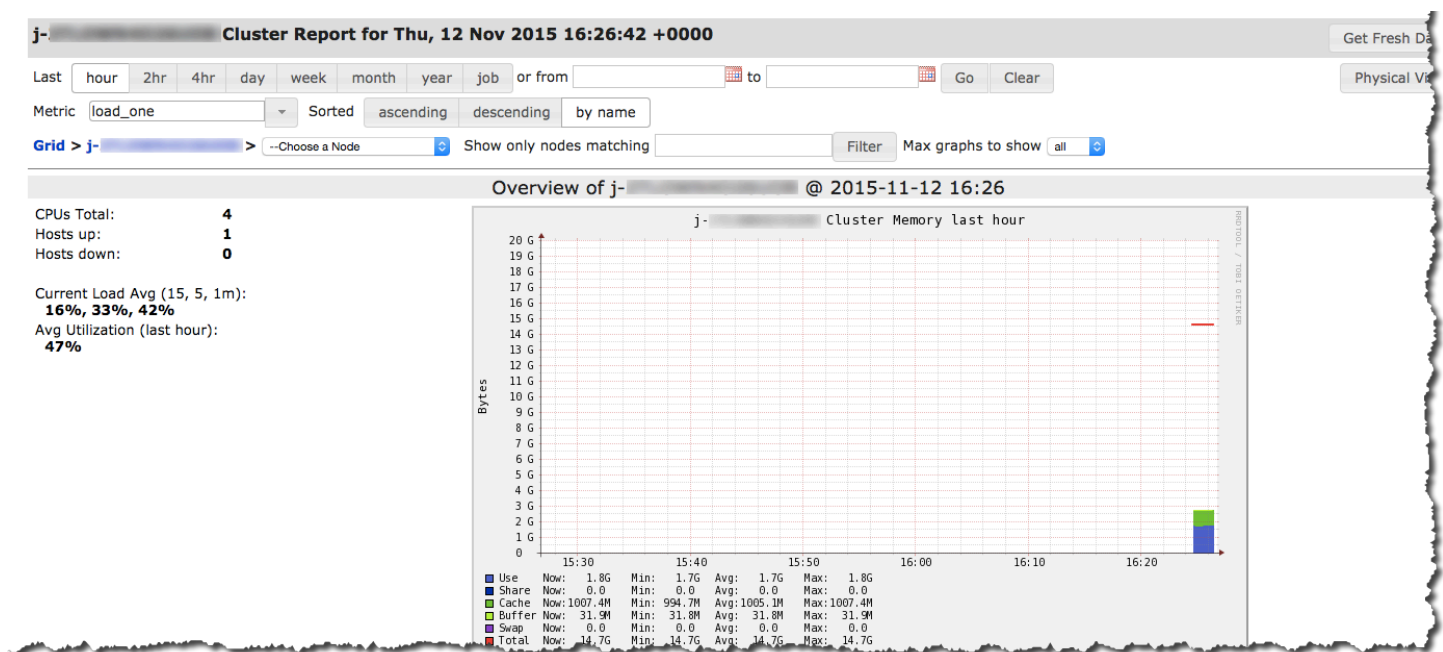
La dernière version d'Amazon EMR à inclure Ganglia était Amazon EMR 6.15.0. Pour surveiller votre cluster, les versions supérieures à 6.15.0 incluent l'agent [Amazon CloudWatch](#).

Ganglia fournit une interface utilisateur basée sur le web que vous pouvez utiliser pour afficher les métriques collectées par Ganglia. Lorsque vous exécutez Ganglia sur Amazon EMR, l'interface web s'exécute sur le nœud principal et peut être affichée à l'aide du réacheminement de port, qualifié également de création d'un tunnel SSH. Pour plus d'informations sur l'affichage des interfaces web sur Amazon EMR, consultez la section [Affichage des interfaces web hébergées sur les clusters EMR](#) dans le Guide de gestion Amazon EMR.

Pour afficher l'interface web Ganglia

1. Utilisez SSH pour accéder par tunnel au nœud maître et créer une connexion sécurisée. Pour plus d'informations sur la façon de créer un tunnel SSH pour le nœud principal,

- consultez [Option 2, partie 1 : Configuration d'un tunnel SSH vers le nœud principal à l'aide du réacheminement de port dynamique](#) dans le Guide de gestion d'Amazon EMR.
- Installez un navigateur Web avec un outil proxy, tel que le FoxyProxy plug-in pour Firefox, afin de créer un proxy SOCKS pour les domaines du type `*ec2*.amazonaws.com*`. Pour plus d'informations, consultez [Option 2, partie 2 : Configuration des paramètres de proxy pour afficher les sites web hébergés sur le nœud principal](#) dans le Guide de gestion d'Amazon EMR.
 - Une fois le proxy défini et la connexion SSH ouverte, vous pouvez afficher l'interface utilisateur de Ganglia en ouvrant une fenêtre de navigateur avec `http://master-public-dns-name/ganglia/`, où se *master-public-dns-name* trouve l'adresse DNS publique du serveur principal dans le cluster EMR.



Métriques Hadoop et Spark dans Ganglia

Note

La dernière version d'Amazon EMR à inclure Ganglia était Amazon EMR 6.15.0. Pour surveiller votre cluster, les versions supérieures à 6.15.0 incluent l'agent [Amazon CloudWatch](#).

Ganglia fournit des rapports sur les métriques Hadoop pour chaque instance. Les différents types de métriques sont préfixés par catégorie : système de fichiers distribué (dfs.*), machine virtuelle Java (jvm.*), (mapred.*) et appels de procédure MapReduce distante (rpc.*).

Les métriques Ganglia basées sur YARN, telles que Spark et Hadoop, ne sont pas disponibles pour les versions EMR 4.4.0 et 4.5.0. Utilisez une version ultérieure pour employer ces métriques.

Les métriques de Ganglia pour Spark ont généralement des préfixes pour l'ID d'application YARN et Spark DAGScheduler. Ainsi, les préfixes suivent ce format :

- DAGScheduler.*
- application_XXXXXXXXXX_XXXX.driver.*
- application_XXXXXXXXXX_XXXX.executor.*

Historique des versions de Ganglia

Note

La dernière version d'Amazon EMR à inclure Ganglia était Amazon EMR 6.15.0. Pour surveiller votre cluster, les versions supérieures à 6.15.0 incluent l'agent [Amazon CloudWatch](#).

Le tableau suivant répertorie la version de Ganglia incluse dans chaque version d'Amazon EMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations de version de Ganglia

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.36.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode,

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.14.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-6.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.12.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.11.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-6.11.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.10.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.10.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-6.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.9.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-6.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.36.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.36.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-6.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.35.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-6.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.3.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-6.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.2.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.2.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-6.1.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.1.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-6.0.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-6.0.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.34.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.33.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.33.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.32.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.32.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.31.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.31.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.30.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.30.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.30.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
mr-5.29.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.28.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.28.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.27.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.27.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.26.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.25.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.24.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.24.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.23.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.23.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.22.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.21.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.21.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.21.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.20.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.20.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.19.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.19.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.18.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.18.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.17.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.17.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.17.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.16.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.16.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.15.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.15.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.14.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.14.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.14.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.13.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.13.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.12.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.12.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.11.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.11.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.11.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.10.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.10.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.9.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.8.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.8.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.7.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-5.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.6.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, webserver
emr-5.5.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.5.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.5.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.5.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.4.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.3.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.3.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.2.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.2.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.2.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.2.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.1.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-5.1.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.0.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.0.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-5.0.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-5.0.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.9.6	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.9.5	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-4.9.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.9.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.9.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-4.9.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.8.5	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.8.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-4.8.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.8.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.8.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-4.8.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.7.4	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver
emr-4.7.3	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-4.7.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.7.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.7.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-4.6.1	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.6.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.5.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Étiquette de version Amazon EMR	Version Ganglia	Composants installés avec Ganglia
emr-4.4.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.3.0	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver
emr-4.2.0	3.6.0	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, webserver

Apache Hadoop

[Apache Hadoop](#) est une infrastructure logicielle Java open source qui prend en charge le traitement d'importants volumes de données sur un cluster d'instances. Il peut s'exécuter sur une seule instance ou sur des milliers d'instances. Hadoop utilise différents modèles de traitement, tels que MapReduce Tez, pour répartir le traitement sur plusieurs instances et utilise également un système de fichiers distribué appelé HDFS pour stocker les données sur plusieurs instances. Hadoop surveille l'intégrité des instances dans le cluster et peut se remettre d'une défaillance d'un ou de plusieurs nœuds. De cette manière, Hadoop fournit une capacité accrue de stockage et de traitement, ainsi qu'une disponibilité élevée. Pour plus d'informations, consultez la [documentation Hadoop](#).

Le tableau suivant répertorie la version de Hadoop incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Hadoop.

Pour connaître la version des composants installés avec Hadoop dans cette version, consultez la version [7.2.0 des versions](#) des composants.

Informations sur la version de Hadoop pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version de Hadoop	Composants installés avec Hadoop
emr-7.2.0	Hadoop 3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Le tableau suivant répertorie la version de Hadoop incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Hadoop.

Pour connaître la version des composants installés avec Hadoop dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Hadoop pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version de Hadoop	Composants installés avec Hadoop
emr-6.15.0	Hadoop 3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Le tableau suivant répertorie la version de Hadoop incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Hadoop.

Pour la version des composants installés avec Hadoop dans cette version, consultez la [version 5.36.2 Versions des composants](#).

Informations sur la version de Hadoop pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version de Hadoop	Composants installés avec Hadoop
emr-5.36.2	Hadoop 2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager

Étiquette EMR de sortie Amazon	Version de Hadoop	Composants installés avec Hadoop
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

À partir d'Amazon EMR 5.18.0, vous pouvez utiliser le référentiel d'EMRartefacts Amazon pour créer votre code de travail en fonction des versions exactes des bibliothèques et des dépendances disponibles avec des versions spécifiques d'Amazon. EMR Pour de plus amples informations, veuillez consulter [Vérification des dépendances à l'aide du référentiel d'artefacts d'Amazon EMR](#).

Rubriques

- [Configuration de Hadoop](#)
- [Chiffrement transparent HDFS sur Amazon EMR](#)
- [Création ou exécution d'une application Hadoop](#)
- [Lire les objets restaurés](#)
- [Activer la détection non uniforme de l'accès à la mémoire pour les conteneurs YARN](#)
- [Historique des versions de Hadoop](#)

Configuration de Hadoop

Les sections suivantes présentent les paramètres de configuration par défaut pour les démons, les tâches et. HDFS

Rubriques

- [Configuration de la tâche](#)
- [Paramètres de configuration de démon Hadoop](#)
- [HDFSconfiguration](#)

Configuration de la tâche

Vous pouvez définir des variables de configuration pour optimiser les performances de vos MapReduce tâches. Cette section fournit les valeurs par défaut des paramètres importants. Les

valeurs par défaut varient en fonction du type d'EC2instance du nœud utilisé dans le cluster. HBaseest disponible lors de l'utilisation de la EMR version 4.6.0 ou ultérieure d'Amazon. Différents paramètres par défaut sont utilisés lors de HBase l'installation. Ces valeurs sont fournies avec les valeurs par défaut initiales.

Hadoop 2 utilise deux paramètres : `mapreduce.map.java.opts` et, pour configurer la mémoire `mapreduce.reduce.java.opts`, respectivement pour cartographier et réduire JVMs. Ces paramètres remplacent l'option de configuration unique `mapreduce.map.java.opts` des versions antérieures de Hadoop.

De la même façon, `mapred.job.jvm.num.tasks` remplace `mapred.job.reuse.jvm.num.tasks` dans Hadoop 2.7.2 et versions ultérieures. Amazon EMR définit cette valeur à 20 quel que soit le type d'EC2instance. Vous pouvez remplacer ce paramètre à l'aide de la classification de configuration `mapred-site`. La définition d'une valeur de -1 indique que a JVM peut être réutilisé pour un nombre infini de tâches au sein d'une même tâche, et une valeur de 1 indique qu'une nouvelle tâche JVM est générée pour chaque tâche.

Par exemple, pour définir la valeur de `mapred.job.jvm.num.tasks` sur -1, vous pouvez créer un fichier contenant les éléments suivants :

```
[
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.job.jvm.num.tasks": "-1"
    }
  }
]
```

Lorsque vous utilisez la `create-cluster` commande ou la `modify-instance-groups` commande du AWS CLI, vous pouvez alors référencer le fichier JSON de configuration. Dans l'exemple suivant, le fichier de configuration est enregistré sous la forme `myConfig.json` et stocké dans Amazon S3.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge \  
--instance-count 3 --applications Name=Hadoop --configurations https://  
s3.amazonaws.com/mybucket/myfolder/myConfig.json \  
--use-default-roles
```

Vous pouvez modifier les valeurs par défaut répertoriées ci-dessous à l'aide de la classification de `mapred-site` configuration de la même manière, et définir plusieurs valeurs et plusieurs classifications de configuration à l'aide d'un seul JSON fichier. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier des classifications de configuration supplémentaires pour chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, utilisez la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).

Valeurs par défaut des paramètres de configuration des tâches

Types d'instances

- [Instances c1](#)
- [Instances c3](#)
- [Instances c4](#)
- [Instances c5](#)
- [Instances c5a](#)
- [Instances c5ad](#)
- [Instances c5d](#)
- [Instances c5n](#)
- [Instances c6a](#)
- [instances c6g](#)
- [instances c6gd](#)
- [Instances c6gn](#)
- [Instances c6i](#)
- [Instances c6id](#)

- [Instances c6in](#)
- [instances C7a](#)
- [instances c7g](#)
- [Instances c7gd](#)
- [Instances c7gn](#)
- [Instances c7i](#)
- [Instances d2](#)
- [Instances d3](#)
- [instances d3en](#)
- [Instances g3](#)
- [Instances g3s](#)
- [Instances g4dn](#)
- [Instances g5](#)
- [Instances h1](#)
- [Instances i2](#)
- [Instances i3](#)
- [Instances i3en](#)
- [Instances i4g](#)
- [Instances i4i](#)
- [Instances im4gn](#)
- [Instances is4gen](#)
- [Instances m1](#)
- [Instances m2](#)
- [Instances m3](#)
- [Instances m4](#)
- [Instances m5](#)
- [Instances m5a](#)
- [instances m5ad](#)
- [Instances m5d](#)

- [instances m5dn](#)
- [instances m5n](#)
- [instances m5zn](#)
- [Instances m6a](#)
- [Instances m6g](#)
- [instances m6gd](#)
- [Instances m6i](#)
- [Instances m6id](#)
- [instances m6idn](#)
- [instances m6in](#)
- [Instances m7a](#)
- [Instances m7g](#)
- [Instances m7gd](#)
- [Instances m7i](#)
- [Instances m7i-flex](#)
- [Instances p2](#)
- [Instances p3](#)
- [Instances p5](#)
- [Instances r3](#)
- [Instances r4](#)
- [Instances r5](#)
- [Instances r5a](#)
- [instances r5ad](#)
- [instances r5b](#)
- [Instances r5d](#)
- [instances r5dn](#)
- [Instances r5n](#)
- [Instances r6a](#)
- [instances r6g](#)
- [Instances r6gd](#)

- [Instances r6i](#)
- [Instances r6id](#)
- [instances r6idn](#)
- [Instances r6in](#)
- [Instances r7a](#)
- [instances r7g](#)
- [Instances r7gd](#)
- [Instances R7i](#)
- [Instances r7iz](#)
- [Instances x1](#)
- [Instances x1e](#)
- [Instances x2gd](#)
- [instances x2idn](#)
- [Instances x2iedn](#)
- [instances z1d](#)

Instances c1

c1.medium

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>
<code>mapreduce.map.memory.mb</code>	512	512
<code>mapreduce.reduce.memory.mb</code>	512	512
<code>yarn.app.mapreduce.am.resource.mb</code>	512	512
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.maximum-allocation-mb</code>	512	512
<code>yarn.nodemanager.resource-memory-mb</code>	1 024	512

c1.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx864m</code>	<code>-Xmx864m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1536m</code>	<code>-Xmx1536m</code>
<code>mapreduce.map.memory.mb</code>	1 024	1 024
<code>mapreduce.reduce.memory.mb</code>	2048	2048
<code>yarn.app.mapreduce.am.resource.mb</code>	2048	2048
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	2048	2560
<code>yarn.nodemanager.resource-memory-mb</code>	5120	2560

Instances c3

c3.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1126m</code>	<code>-Xmx1126m</code>

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.java.opts</code>	<code>-Xmx2252m</code>	<code>-Xmx2252m</code>
<code>mapreduce.map.memory.mb</code>	1 408	1 408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c3.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

c3.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23040	11520
<code>yarn.nodemanager.resource.memory-mb</code>	23040	11520

c3.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1331m</code>	<code>-Xmx1331m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2662m</code>	<code>-Xmx2662m</code>
<code>mapreduce.map.memory.mb</code>	1664	1664

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.reduce.memory.mb</code>	3328	3328
<code>yarn.app.mapreduce.am.resource.mb</code>	3328	3328
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	53248	26624
<code>yarn.nodemanager.resource.memory-mb</code>	53248	26624

Instances c4

c4.large

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx717m</code>	<code>-Xmx717m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1434m</code>	<code>-Xmx1434m</code>
<code>mapreduce.map.memory.mb</code>	896	896
<code>mapreduce.reduce.memory.mb</code>	1792	1792
<code>yarn.app.mapreduce.am.resource.mb</code>	1792	1792
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1792	896

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.nodemanager.resource.memory-mb</code>	1792	896

c4.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1126m</code>	<code>-Xmx1126m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2252m</code>	<code>-Xmx2252m</code>
<code>mapreduce.map.memory.mb</code>	1 408	1 408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c4.2.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

c4.4.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23040	11520
<code>yarn.nodemanager.resource.memory-mb</code>	23040	11520

c4.8.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1183m	-Xmx1183m
mapreduce.java.opts	-Xmx2366m	-Xmx2366m
mapreduce.map.memory.mb	1479	1479
mapreduce.reduce.memory.mb	2958	2958
yarn.app.mapreduce.am.resource.mb	2958	2958
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	53248	26624
yarn.nodemanager.resource.memory-mb	53248	26624

Instances c5

c5.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	6144	3072
<code>yarn.nodemanager.resource.memory-mb</code>	6144	3072

c5.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

c5.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

c5.9xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1456m</code>	<code>-Xmx1456m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2912m</code>	<code>-Xmx2912m</code>
<code>mapreduce.map.memory.mb</code>	1820	1820
<code>mapreduce.reduce.memory.mb</code>	3640	3640
<code>yarn.app.mapreduce.am.resource.mb</code>	3640	3640

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	65536	32768
<code>yarn.nodemanager.resource.memory-mb</code>	65536	32768

c5.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1502m</code>	<code>-Xmx1502m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3004m</code>	<code>-Xmx3004m</code>
<code>mapreduce.map.memory.mb</code>	1877	1877
<code>mapreduce.reduce.memory.mb</code>	3754	3754
<code>yarn.app.mapreduce.am.resource.mb</code>	3754	3754
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30048
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30048

c5.18xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1547m</code>	<code>-Xmx1547m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3094m</code>	<code>-Xmx3094m</code>
<code>mapreduce.map.memory.mb</code>	1934	1934
<code>mapreduce.reduce.memory.mb</code>	3868	3868
<code>yarn.app.mapreduce.am.resource.mb</code>	3868	3868
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	139264	30960
<code>yarn.nodemanager.resource.memory-mb</code>	139264	30960

c5.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1570m</code>	<code>-Xmx1570m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.map.memory.mb</code>	1963	1963
<code>mapreduce.reduce.memory.mb</code>	3926	3926
<code>yarn.app.mapreduce.am.resource.mb</code>	3926	3926

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31376
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31376

Instances c5a

c5a.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1126m</code>	<code>-Xmx1126m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2252m</code>	<code>-Xmx2252m</code>
<code>mapreduce.map.memory.mb</code>	1 408	1 408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c5a.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c5a.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c5a.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c5a.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1502m	-Xmx1502m
mapreduce.java.opts	-Xmx3004m	-Xmx3004m
mapreduce.map.memory.mb	1877	1877
mapreduce.reduce.memory.mb	3754	3754
yarn.app.mapreduce.am.resource.mb	3754	3754
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	90112	30048
yarn.nodemanager.resource.memory-mb	90112	30048

c5a.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

c5a.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1494m</code>	<code>-Xmx1494m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2988m</code>	<code>-Xmx2988m</code>
<code>mapreduce.map.memory.mb</code>	1867	1867
<code>mapreduce.reduce.memory.mb</code>	3734	3734
<code>yarn.app.mapreduce.am.resource.mb</code>	3734	3734
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

Instances c5ad

c5ad.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1126m</code>	<code>-Xmx1126m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2252m</code>	<code>-Xmx2252m</code>
<code>mapreduce.map.memory.mb</code>	1 408	1 408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c5ad.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c5ad.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c5ad.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c5ad.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource-memory-mb</code>	85504	32074

c5ad.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource-memory-mb</code>	116736	29184

c5ad.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

Instances c5d

c5d.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1229m	-Xmx1229m
mapreduce.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.map.memory.mb	1536	1536
mapreduce.reduce.memory.mb	3072	3072
yarn.app.mapreduce.am.resource.mb	3072	3072

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	6144	3072
<code>yarn.nodemanager.resource.memory-mb</code>	6144	3072

c5d.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

c5d.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

c5d.9xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1456m</code>	<code>-Xmx1456m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2912m</code>	<code>-Xmx2912m</code>
<code>mapreduce.map.memory.mb</code>	1820	1820
<code>mapreduce.reduce.memory.mb</code>	3640	3640
<code>yarn.app.mapreduce.am.resource.mb</code>	3640	3640

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	65536	32768
<code>yarn.nodemanager.resource.memory-mb</code>	65536	32768

c5d.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1502m</code>	<code>-Xmx1502m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3004m</code>	<code>-Xmx3004m</code>
<code>mapreduce.map.memory.mb</code>	1877	1877
<code>mapreduce.reduce.memory.mb</code>	3754	3754
<code>yarn.app.mapreduce.am.resource.mb</code>	3754	3754
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30048
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30048

c5d.18xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1547m</code>	<code>-Xmx1547m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3094m</code>	<code>-Xmx3094m</code>
<code>mapreduce.map.memory.mb</code>	1934	1934
<code>mapreduce.reduce.memory.mb</code>	3868	3868
<code>yarn.app.mapreduce.am.resource.mb</code>	3868	3868
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	139264	30960
<code>yarn.nodemanager.resource.memory-mb</code>	139264	30960

c5d.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1570m</code>	<code>-Xmx1570m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.map.memory.mb</code>	1963	1963
<code>mapreduce.reduce.memory.mb</code>	3926	3926
<code>yarn.app.mapreduce.am.resource.mb</code>	3926	3926

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31376
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31376

Instances c5n

c5n.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1613m</code>	<code>-Xmx1613m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3226m</code>	<code>-Xmx3226m</code>
<code>mapreduce.map.memory.mb</code>	2016	2016
<code>mapreduce.reduce.memory.mb</code>	4032	4032
<code>yarn.app.mapreduce.am.resource.mb</code>	4032	4032
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	8064	4032
<code>yarn.nodemanager.resource.memory-mb</code>	8064	4032

c5n.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1613m</code>	<code>-Xmx1613m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3226m</code>	<code>-Xmx3226m</code>
<code>mapreduce.map.memory.mb</code>	2016	2016
<code>mapreduce.reduce.memory.mb</code>	4032	4032
<code>yarn.app.mapreduce.am.resource.mb</code>	4032	4032
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	16128	8064
<code>yarn.nodemanager.resource.memory-mb</code>	16128	8064

c5n.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1741m</code>	<code>-Xmx1741m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3482m</code>	<code>-Xmx3482m</code>
<code>mapreduce.map.memory.mb</code>	2176	2176
<code>mapreduce.reduce.memory.mb</code>	4352	4352
<code>yarn.app.mapreduce.am.resource.mb</code>	4352	4352

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	34816	17408
<code>yarn.nodemanager.resource.memory-mb</code>	34816	17408

c5n.9xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2002m</code>	<code>-Xmx2002m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4004m</code>	<code>-Xmx4004m</code>
<code>mapreduce.map.memory.mb</code>	2503	2503
<code>mapreduce.reduce.memory.mb</code>	5006	5006
<code>yarn.app.mapreduce.am.resource.mb</code>	5006	5006
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30040
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30040

c5n.18xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2094m	-Xmx2094m
mapreduce.java.opts	-Xmx4188m	-Xmx4188m
mapreduce.map.memory.mb	2617	2617
mapreduce.reduce.memory.mb	5234	5234
yarn.app.mapreduce.am.resource.mb	5234	5234
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31396
yarn.nodemanager.resource.memory-mb	188416	31396

Instances c6a

c6a.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6a.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c6a.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6a.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c6a.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	-Xmx1425m	-Xmx1425m
<code>mapreduce.java.opts</code>	-Xmx2850m	-Xmx2850m
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6a.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

c6a.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1494m</code>	<code>-Xmx1494m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2988m</code>	<code>-Xmx2988m</code>
<code>mapreduce.map.memory.mb</code>	1867	1867
<code>mapreduce.reduce.memory.mb</code>	3734	3734
<code>yarn.app.mapreduce.am.resource.mb</code>	3734	3734

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

c6a.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1510m</code>	<code>-Xmx1510m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3020m</code>	<code>-Xmx3020m</code>
<code>mapreduce.map.memory.mb</code>	1888	1888
<code>mapreduce.reduce.memory.mb</code>	3776	3776
<code>yarn.app.mapreduce.am.resource.mb</code>	3776	3776
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

c6a.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1527m	-Xmx1527m
mapreduce.java.opts	-Xmx3054m	-Xmx3054m
mapreduce.map.memory.mb	1909	1909
mapreduce.reduce.memory.mb	3818	3818
yarn.app.mapreduce.am.resource.mb	3818	3818
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30608
yarn.nodemanager.resource.memory-mb	366592	30608

instances c6g

c6g.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6g.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c6g.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c6g.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c6g.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6g.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

instances c6gd

c6gd.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6gd.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c6gd.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6gd.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c6gd.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6gd.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

Instances c6gn

c6gn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6gn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c6gn.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1171m	-Xmx1171m
mapreduce.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.map.memory.mb	1464	1464
mapreduce.reduce.memory.mb	2928	2928
yarn.app.mapreduce.am.resource.mb	2928	2928
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

c6gn.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c6gn.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6gn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

Instances c6i

c6i.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6i.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c6i.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c6i.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c6i.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6i.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c6i.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

c6i.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1510m</code>	<code>-Xmx1510m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3020m</code>	<code>-Xmx3020m</code>
<code>mapreduce.map.memory.mb</code>	1888	1888
<code>mapreduce.reduce.memory.mb</code>	3776	3776
<code>yarn.app.mapreduce.am.resource.mb</code>	3776	3776
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Instances c6id

c6id.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1126m</code>	<code>-Xmx1126m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2252m</code>	<code>-Xmx2252m</code>
<code>mapreduce.map.memory.mb</code>	1 408	1 408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6id.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c6id.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c6id.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1357m	-Xmx1357m
mapreduce.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.map.memory.mb	1696	1696
mapreduce.reduce.memory.mb	3392	3392
yarn.app.mapreduce.am.resource.mb	3392	3392
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

c6id.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c6id.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1459m</code>	<code>-Xmx1459m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.map.memory.mb</code>	1824	1824
<code>mapreduce.reduce.memory.mb</code>	3648	3648
<code>yarn.app.mapreduce.am.resource.mb</code>	3648	3648
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

c6id.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29840
yarn.nodemanager.resource.memory-mb	179200	29840

c6id.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Instances c6in

c6in.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1126m</code>	<code>-Xmx1126m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2252m</code>	<code>-Xmx2252m</code>
<code>mapreduce.map.memory.mb</code>	1 408	1 408
<code>mapreduce.reduce.memory.mb</code>	2816	2816
<code>yarn.app.mapreduce.am.resource.mb</code>	2816	2816
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c6in.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c6in.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c6in.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c6in.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1425m	-Xmx1425m
mapreduce.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.map.memory.mb	1781	1781
mapreduce.reduce.memory.mb	3562	3562
yarn.app.mapreduce.am.resource.mb	3562	3562
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	85504	32074
yarn.nodemanager.resource.memory-mb	85504	32074

c6in.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

c6in.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1494m</code>	<code>-Xmx1494m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2988m</code>	<code>-Xmx2988m</code>
<code>mapreduce.map.memory.mb</code>	1867	1867
<code>mapreduce.reduce.memory.mb</code>	3734	3734
<code>yarn.app.mapreduce.am.resource.mb</code>	3734	3734
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

c6in.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1510m	-Xmx1510m
mapreduce.java.opts	-Xmx3020m	-Xmx3020m
mapreduce.map.memory.mb	1888	1888
mapreduce.reduce.memory.mb	3776	3776
yarn.app.mapreduce.am.resource.mb	3776	3776
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

instances C7a

c7a.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c7a.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c7a.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c7a.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c7a.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	-Xmx1425m	-Xmx1425m
<code>mapreduce.java.opts</code>	-Xmx2850m	-Xmx2850m
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c7a.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c7a.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

c7a.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1510m</code>	<code>-Xmx1510m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3020m</code>	<code>-Xmx3020m</code>
<code>mapreduce.map.memory.mb</code>	1888	1888
<code>mapreduce.reduce.memory.mb</code>	3776	3776
<code>yarn.app.mapreduce.am.resource.mb</code>	3776	3776
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

c7a.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1527m	-Xmx1527m
mapreduce.java.opts	-Xmx3054m	-Xmx3054m
mapreduce.map.memory.mb	1909	1909
mapreduce.reduce.memory.mb	3818	3818
yarn.app.mapreduce.am.resource.mb	3818	3818
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30608
yarn.nodemanager.resource.memory-mb	366592	30608

instances c7g

c7g.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c7g.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c7g.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c7g.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c7g.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c7g.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

Instances c7gd

c7gd.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c7gd.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c7gd.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c7gd.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c7gd.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c7gd.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

Instances c7gn

c7gn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c7gn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c7gn.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c7gn.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c7gn.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c7gn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

Instances c7i

c7i.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1126m	-Xmx1126m
mapreduce.java.opts	-Xmx2252m	-Xmx2252m
mapreduce.map.memory.mb	1 408	1 408
mapreduce.reduce.memory.mb	2816	2816
yarn.app.mapreduce.am.resource.mb	2816	2816

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	5632	2816
<code>yarn.nodemanager.resource.memory-mb</code>	5632	2816

c7i.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

c7i.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1171m</code>	<code>-Xmx1171m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.map.memory.mb</code>	1464	1464
<code>mapreduce.reduce.memory.mb</code>	2928	2928
<code>yarn.app.mapreduce.am.resource.mb</code>	2928	2928
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

c7i.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1357m</code>	<code>-Xmx1357m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.map.memory.mb</code>	1696	1696
<code>mapreduce.reduce.memory.mb</code>	3392	3392
<code>yarn.app.mapreduce.am.resource.mb</code>	3392	3392

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

c7i.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1425m</code>	<code>-Xmx1425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.map.memory.mb</code>	1781	1781
<code>mapreduce.reduce.memory.mb</code>	3562	3562
<code>yarn.app.mapreduce.am.resource.mb</code>	3562	3562
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32074
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32074

c7i.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1459m	-Xmx1459m
mapreduce.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.map.memory.mb	1824	1824
mapreduce.reduce.memory.mb	3648	3648
yarn.app.mapreduce.am.resource.mb	3648	3648
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

c7i.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx1494m	-Xmx1494m
mapreduce.java.opts	-Xmx2988m	-Xmx2988m
mapreduce.map.memory.mb	1867	1867
mapreduce.reduce.memory.mb	3734	3734
yarn.app.mapreduce.am.resource.mb	3734	3734

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29840
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29840

c7i.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1527m</code>	<code>-Xmx1527m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3054m</code>	<code>-Xmx3054m</code>
<code>mapreduce.map.memory.mb</code>	1909	1909
<code>mapreduce.reduce.memory.mb</code>	3818	3818
<code>yarn.app.mapreduce.am.resource.mb</code>	3818	3818
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30608
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30608

Instances d2

d2.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

d2.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

d2.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

d2.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2417m</code>	<code>-Xmx2417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4834m</code>	<code>-Xmx4834m</code>
<code>mapreduce.map.memory.mb</code>	3021	3021
<code>mapreduce.reduce.memory.mb</code>	6042	6042
<code>yarn.app.mapreduce.am.resource.mb</code>	6042	6042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30194
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30194

Instances d3

d3.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

d3.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

d3.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

d3.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

instances d3en

d3en.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

d3en.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

d3en.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

d3en.6xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2850m</code>	<code>-Xmx2850m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5700m</code>	<code>-Xmx5700m</code>
<code>mapreduce.map.memory.mb</code>	3563	3563
<code>mapreduce.reduce.memory.mb</code>	7126	7126
<code>yarn.app.mapreduce.am.resource.mb</code>	7126	7126
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	28496
<code>yarn.nodemanager.resource.memory-mb</code>	85504	28496

d3en.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

d3en.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

Instances g3

g3.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

g3.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

g3.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances g3s

g3s.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

Instances g4dn

g4dn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

g4dn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

g4dn.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2867m</code>	<code>-Xmx2867m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.map.memory.mb</code>	3584	3584
<code>mapreduce.reduce.memory.mb</code>	7168	7168
<code>yarn.app.mapreduce.am.resource.mb</code>	7168	7168
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

g4dn.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

g4dn.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.java.opts	-Xmx6280m	-Xmx6280m
mapreduce.map.memory.mb	3925	3925
mapreduce.reduce.memory.mb	7850	7850
yarn.app.mapreduce.am.resource.mb	7850	7850

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31416
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31416

g4dn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3174m</code>	<code>-Xmx3174m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6348m</code>	<code>-Xmx6348m</code>
<code>mapreduce.map.memory.mb</code>	3 968	3 968
<code>mapreduce.reduce.memory.mb</code>	7936	7936
<code>yarn.app.mapreduce.am.resource.mb</code>	7936	7936
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253 952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253 952	31744

Instances g5

g5.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

g5.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

g5.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

g5.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

g5.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

g5.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

g5.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

g5.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3089m</code>	<code>-Xmx3089m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.map.memory.mb</code>	3861	3861
<code>mapreduce.reduce.memory.mb</code>	7722	7722
<code>yarn.app.mapreduce.am.resource.mb</code>	7722	7722

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30952
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30952

Instances h1

h1.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

h1.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2867m</code>	<code>-Xmx2867m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.map.memory.mb</code>	3584	3584
<code>mapreduce.reduce.memory.mb</code>	7168	7168
<code>yarn.app.mapreduce.am.resource.mb</code>	7168	7168
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

h1.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource-memory-mb</code>	122880	30720

h1.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3174m</code>	<code>-Xmx3174m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6348m</code>	<code>-Xmx6348m</code>
<code>mapreduce.map.memory.mb</code>	3 968	3 968
<code>mapreduce.reduce.memory.mb</code>	7936	7936
<code>yarn.app.mapreduce.am.resource.mb</code>	7936	7936
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253 952	31744
<code>yarn.nodemanager.resource-memory-mb</code>	253 952	31744

Instances i2

i2.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

i2.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

i2.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

i2.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Instances i3

i3.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

i3.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

i3.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

i3.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

i3.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances i3en

i3en.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4915m</code>	<code>-Xmx4915m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9830m</code>	<code>-Xmx9830m</code>
<code>mapreduce.map.memory.mb</code>	6144	6144
<code>mapreduce.reduce.memory.mb</code>	12288	12288
<code>yarn.app.mapreduce.am.resource.mb</code>	12288	12288
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

i3en.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11468m</code>	<code>-Xmx11468m</code>
<code>mapreduce.map.memory.mb</code>	7168	7168
<code>mapreduce.reduce.memory.mb</code>	14336	14336
<code>yarn.app.mapreduce.am.resource.mb</code>	14336	14336

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

i3en.3xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6007m</code>	<code>-Xmx6007m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12014m</code>	<code>-Xmx12014m</code>
<code>mapreduce.map.memory.mb</code>	7509	7509
<code>mapreduce.reduce.memory.mb</code>	15018	15018
<code>yarn.app.mapreduce.am.resource.mb</code>	15018	15018
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30040
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30040

i3en.6xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6281m</code>	<code>-Xmx6281m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12562m</code>	<code>-Xmx12562m</code>
<code>mapreduce.map.memory.mb</code>	7851	7851
<code>mapreduce.reduce.memory.mb</code>	15702	15702
<code>yarn.app.mapreduce.am.resource.mb</code>	15702	15702
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31396
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31396

i3en.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

i3en.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6486m</code>	<code>-Xmx6486m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12972m</code>	<code>-Xmx12972m</code>
<code>mapreduce.map.memory.mb</code>	8107	8107
<code>mapreduce.reduce.memory.mb</code>	16214	16214
<code>yarn.app.mapreduce.am.resource.mb</code>	16214	16214
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	778240	32396
<code>yarn.nodemanager.resource.memory-mb</code>	778240	32396

Instances i4g

i4g.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

i4g.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

i4g.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

i4g.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

i4g.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances i4i

i4i.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

i4i.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

i4i.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

i4i.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

i4i.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

i4i.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

i4i.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

i4i.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

Instances im4gn

im4gn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

im4gn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

im4gn.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

im4gn.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

im4gn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Instances is4gen

is4gen.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3514m</code>	<code>-Xmx3514m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx7028m</code>	<code>-Xmx7028m</code>
<code>mapreduce.map.memory.mb</code>	4393	4393
<code>mapreduce.reduce.memory.mb</code>	8786	8786
<code>yarn.app.mapreduce.am.resource.mb</code>	8786	8786
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	17572,12	8786,06
<code>yarn.nodemanager.resource.memory-mb</code>	17572,12	8786,06

is4gen.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3866m</code>	<code>-Xmx3866m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx7732m</code>	<code>-Xmx7732m</code>
<code>mapreduce.map.memory.mb</code>	4832	4832
<code>mapreduce.reduce.memory.mb</code>	9664	9664
<code>yarn.app.mapreduce.am.resource.mb</code>	9664	9664

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	38656	19328
<code>yarn.nodemanager.resource.memory-mb</code>	38656	19328

is4gen.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4275m</code>	<code>-Xmx4275m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx8550m</code>	<code>-Xmx8550m</code>
<code>mapreduce.map.memory.mb</code>	5344	5344
<code>mapreduce.reduce.memory.mb</code>	10688	10688
<code>yarn.app.mapreduce.am.resource.mb</code>	10688	10688
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	32064
<code>yarn.nodemanager.resource.memory-mb</code>	85504	32064

is4gen.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4480m</code>	<code>-Xmx4480m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx8960m</code>	<code>-Xmx8960m</code>
<code>mapreduce.map.memory.mb</code>	5600	5600
<code>mapreduce.reduce.memory.mb</code>	11200	11200
<code>yarn.app.mapreduce.am.resource.mb</code>	11200	11200
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	22400
<code>yarn.nodemanager.resource.memory-mb</code>	179200	22400

Instances m1

m1.small

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx288m</code>	<code>-Xmx288m</code>
<code>mapreduce.map.memory.mb</code>	512	512
<code>mapreduce.reduce.memory.mb</code>	512	512
<code>yarn.app.mapreduce.am.resource.mb</code>	512	512

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	512	512
<code>yarn.nodemanager.resource.memory-mb</code>	1 024	512

m1.medium

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx512m</code>	<code>-Xmx512m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx768m</code>	<code>-Xmx768m</code>
<code>mapreduce.map.memory.mb</code>	768	768
<code>mapreduce.reduce.memory.mb</code>	1 024	1 024
<code>yarn.app.mapreduce.am.resource.mb</code>	1 024	1 024
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	2048	1 024
<code>yarn.nodemanager.resource.memory-mb</code>	2048	1 024

m1.large

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx512m	-Xmx512m
mapreduce.java.opts	-Xmx1024m	-Xmx1024m
mapreduce.map.memory.mb	768	768
mapreduce.reduce.memory.mb	1536	1536
yarn.app.mapreduce.am.resource.mb	1536	1536
yarn.scheduler.minimum-allocation-mb	256	256
yarn.scheduler.maximum-allocation-mb	3072	2560
yarn.nodemanager.resource.memory-mb	5120	2560

m1.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx512m	-Xmx512m
mapreduce.java.opts	-Xmx1536m	-Xmx1536m
mapreduce.map.memory.mb	768	768
mapreduce.reduce.memory.mb	2048	2048
yarn.app.mapreduce.am.resource.mb	2048	2048

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	8192	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

Instances m2

m2.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx864m</code>	<code>-Xmx864m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx1536m</code>	<code>-Xmx1536m</code>
<code>mapreduce.map.memory.mb</code>	1 024	1 024
<code>mapreduce.reduce.memory.mb</code>	2048	2048
<code>yarn.app.mapreduce.am.resource.mb</code>	2048	2048
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	7168	7168
<code>yarn.nodemanager.resource.memory-mb</code>	14336	7168

m2.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1280m</code>	<code>-Xmx1280m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	2560	2560
<code>yarn.app.mapreduce.am.resource.mb</code>	2560	2560
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	8192	15360
<code>yarn.nodemanager.resource.memory-mb</code>	30720	15360

m2.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1280m</code>	<code>-Xmx1280m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	2560	2560
<code>yarn.app.mapreduce.am.resource.mb</code>	2560	2560

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	8192	30720
<code>yarn.nodemanager.resource.memory-mb</code>	61440	30720

Instances m3

m3.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

m3.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1152m</code>	<code>-Xmx1152m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2304m</code>	<code>-Xmx2304m</code>
<code>mapreduce.map.memory.mb</code>	1440	1440
<code>mapreduce.reduce.memory.mb</code>	2880	2880
<code>yarn.app.mapreduce.am.resource.mb</code>	2880	2880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23040	11520
<code>yarn.nodemanager.resource.memory-mb</code>	23040	11520

Instances m4

m4.large

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	6144	3072
<code>yarn.nodemanager.resource.memory-mb</code>	6144	3072

m4.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

m4.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1229m</code>	<code>-Xmx1229m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.map.memory.mb</code>	1536	1536
<code>mapreduce.reduce.memory.mb</code>	3072	3072
<code>yarn.app.mapreduce.am.resource.mb</code>	3072	3072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

m4.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1434m</code>	<code>-Xmx1434m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2868m</code>	<code>-Xmx2868m</code>
<code>mapreduce.map.memory.mb</code>	1792	1792
<code>mapreduce.reduce.memory.mb</code>	3584	3584
<code>yarn.app.mapreduce.am.resource.mb</code>	3584	3584

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource-memory-mb</code>	57344	28672

m4.10xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1557m</code>	<code>-Xmx1557m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3114m</code>	<code>-Xmx3114m</code>
<code>mapreduce.map.memory.mb</code>	1946	1946
<code>mapreduce.reduce.memory.mb</code>	3892	3892
<code>yarn.app.mapreduce.am.resource.mb</code>	3892	3892
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	155648	31104
<code>yarn.nodemanager.resource-memory-mb</code>	155648	31104

m4.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx1587m</code>	<code>-Xmx1587m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx3174m</code>	<code>-Xmx3174m</code>
<code>mapreduce.map.memory.mb</code>	1 984	1 984
<code>mapreduce.reduce.memory.mb</code>	3 968	3 968
<code>yarn.app.mapreduce.am.resource.mb</code>	3 968	3 968
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253 952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253 952	31744

Instances m5

m5.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource-memory-mb</code>	12288	6144

m5.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource-memory-mb</code>	24576	12288

m5.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2867m</code>	<code>-Xmx2867m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.map.memory.mb</code>	3584	3584
<code>mapreduce.reduce.memory.mb</code>	7168	7168
<code>yarn.app.mapreduce.am.resource.mb</code>	7168	7168
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

m5.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

m5.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6280m</code>	<code>-Xmx6280m</code>
<code>mapreduce.map.memory.mb</code>	3925	3925
<code>mapreduce.reduce.memory.mb</code>	7850	7850
<code>yarn.app.mapreduce.am.resource.mb</code>	7850	7850
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31416
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31416

m5.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3174m</code>	<code>-Xmx3174m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6348m</code>	<code>-Xmx6348m</code>
<code>mapreduce.map.memory.mb</code>	3 968	3 968
<code>mapreduce.reduce.memory.mb</code>	7936	7936
<code>yarn.app.mapreduce.am.resource.mb</code>	7936	7936
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253 952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253 952	31744

m5.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3209m</code>	<code>-Xmx3209m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6418m</code>	<code>-Xmx6418m</code>
<code>mapreduce.map.memory.mb</code>	4011	4011
<code>mapreduce.reduce.memory.mb</code>	8022	8022
<code>yarn.app.mapreduce.am.resource.mb</code>	8022	8022

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32056
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32056

Instances m5a

m5a.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

m5a.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

m5a.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

m5a.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

m5a.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6280m</code>	<code>-Xmx6280m</code>
<code>mapreduce.map.memory.mb</code>	3925	3925
<code>mapreduce.reduce.memory.mb</code>	7850	7850
<code>yarn.app.mapreduce.am.resource.mb</code>	7850	7850
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31416
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31416

m5a.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3174m</code>	<code>-Xmx3174m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6348m</code>	<code>-Xmx6348m</code>
<code>mapreduce.map.memory.mb</code>	3 968	3 968
<code>mapreduce.reduce.memory.mb</code>	7936	7936
<code>yarn.app.mapreduce.am.resource.mb</code>	7936	7936

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253 952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253 952	31744

m5a.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3209m</code>	<code>-Xmx3209m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6418m</code>	<code>-Xmx6418m</code>
<code>mapreduce.map.memory.mb</code>	4011	4011
<code>mapreduce.reduce.memory.mb</code>	8022	8022
<code>yarn.app.mapreduce.am.resource.mb</code>	8022	8022
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32056
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32056

instances m5ad

m5ad.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m5ad.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m5ad.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m5ad.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m5ad.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

m5ad.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m5ad.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

Instances m5d

m5d.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2458m	-Xmx2458m
mapreduce.java.opts	-Xmx4916m	-Xmx4916m
mapreduce.map.memory.mb	3072	3072
mapreduce.reduce.memory.mb	6144	6144
yarn.app.mapreduce.am.resource.mb	6144	6144

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	12288	6144
<code>yarn.nodemanager.resource.memory-mb</code>	12288	6144

m5d.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2458m</code>	<code>-Xmx2458m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4916m</code>	<code>-Xmx4916m</code>
<code>mapreduce.map.memory.mb</code>	3072	3072
<code>mapreduce.reduce.memory.mb</code>	6144	6144
<code>yarn.app.mapreduce.am.resource.mb</code>	6144	6144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

m5d.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2867m	-Xmx2867m
mapreduce.java.opts	-Xmx5734m	-Xmx5734m
mapreduce.map.memory.mb	3584	3584
mapreduce.reduce.memory.mb	7168	7168
yarn.app.mapreduce.am.resource.mb	7168	7168
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	57344	28672
yarn.nodemanager.resource.memory-mb	57344	28672

m5d.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

m5d.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3140m</code>	<code>-Xmx3140m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6280m</code>	<code>-Xmx6280m</code>
<code>mapreduce.map.memory.mb</code>	3925	3925
<code>mapreduce.reduce.memory.mb</code>	7850	7850
<code>yarn.app.mapreduce.am.resource.mb</code>	7850	7850
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31416
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31416

m5d.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3 968	3 968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	253 952	31744
yarn.nodemanager.resource.memory-mb	253 952	31744

m5d.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3209m	-Xmx3209m
mapreduce.java.opts	-Xmx6418m	-Xmx6418m
mapreduce.map.memory.mb	4011	4011
mapreduce.reduce.memory.mb	8022	8022
yarn.app.mapreduce.am.resource.mb	8022	8022

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32056
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32056

instances m5dn

m5dn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m5dn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m5dn.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m5dn.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m5dn.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m5dn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m5dn.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

instances m5n

m5n.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m5n.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m5n.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m5n.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m5n.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

m5n.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m5n.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

instances m5zn

m5zn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.java.opts	-Xmx4608m	-Xmx4608m
mapreduce.map.memory.mb	2880	2880
mapreduce.reduce.memory.mb	5760	5760
yarn.app.mapreduce.am.resource.mb	5760	5760

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

m5zn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m5zn.3xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2577m	-Xmx2577m
mapreduce.java.opts	-Xmx5154m	-Xmx5154m
mapreduce.map.memory.mb	3221	3221
mapreduce.reduce.memory.mb	6442	6442
yarn.app.mapreduce.am.resource.mb	6442	6442
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	38656	19328
yarn.nodemanager.resource.memory-mb	38656	19328

m5zn.6xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2850m	-Xmx2850m
mapreduce.java.opts	-Xmx5700m	-Xmx5700m
mapreduce.map.memory.mb	3563	3563
mapreduce.reduce.memory.mb	7126	7126
yarn.app.mapreduce.am.resource.mb	7126	7126

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	85504	28496
<code>yarn.nodemanager.resource.memory-mb</code>	85504	28496

m5zn.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

Instances m6a

m6a.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6a.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6a.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6a.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6a.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

m6a.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m6a.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m6a.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

m6a.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3089m</code>	<code>-Xmx3089m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.map.memory.mb</code>	3861	3861
<code>mapreduce.reduce.memory.mb</code>	7722	7722
<code>yarn.app.mapreduce.am.resource.mb</code>	7722	7722
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30952
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30952

Instances m6g

m6g.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6g.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6g.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6g.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6g.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

m6g.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

instances m6gd

m6gd.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6gd.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6gd.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6gd.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6gd.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

m6gd.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Instances m6i

m6i.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6i.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6i.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6i.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m6i.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

m6i.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m6i.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m6i.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances m6id

m6id.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6id.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m6id.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6id.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m6id.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m6id.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m6id.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

m6id.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

instances m6idn

m6idn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6idn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6idn.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m6idn.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m6idn.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

m6idn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

m6idn.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

m6idn.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3072m</code>	<code>-Xmx3072m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.map.memory.mb</code>	3840	3840
<code>mapreduce.reduce.memory.mb</code>	7680	7680
<code>yarn.app.mapreduce.am.resource.mb</code>	7680	7680
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

instances m6in

m6in.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m6in.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m6in.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m6in.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

m6in.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

m6in.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m6in.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3055m	-Xmx3055m
mapreduce.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.map.memory.mb	3819	3819
mapreduce.reduce.memory.mb	7638	7638
yarn.app.mapreduce.am.resource.mb	7638	7638
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30520
yarn.nodemanager.resource.memory-mb	366592	30520

m6in.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances m7a

m7a.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7a.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m7a.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m7a.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m7a.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2986m</code>	<code>-Xmx2986m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5972m</code>	<code>-Xmx5972m</code>
<code>mapreduce.map.memory.mb</code>	3733	3733
<code>mapreduce.reduce.memory.mb</code>	7466	7466
<code>yarn.app.mapreduce.am.resource.mb</code>	7466	7466
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	29880
<code>yarn.nodemanager.resource.memory-mb</code>	179200	29880

m7a.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m7a.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

m7a.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3072m	-Xmx3072m
mapreduce.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.map.memory.mb	3840	3840
mapreduce.reduce.memory.mb	7680	7680
yarn.app.mapreduce.am.resource.mb	7680	7680
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

m7a.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3089m	-Xmx3089m
mapreduce.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.map.memory.mb	3861	3861
mapreduce.reduce.memory.mb	7722	7722
yarn.app.mapreduce.am.resource.mb	7722	7722

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30952
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30952

Instances m7g

m7g.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7g.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m7g.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m7g.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m7g.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m7g.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Instances m7gd

m7gd.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7gd.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	23424	11712
yarn.nodemanager.resource.memory-mb	23424	11712

m7gd.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m7gd.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m7gd.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2986m	-Xmx2986m
mapreduce.java.opts	-Xmx5972m	-Xmx5972m
mapreduce.map.memory.mb	3733	3733
mapreduce.reduce.memory.mb	7466	7466
yarn.app.mapreduce.am.resource.mb	7466	7466
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	29880
yarn.nodemanager.resource.memory-mb	179200	29880

m7gd.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3021m	-Xmx3021m
mapreduce.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.map.memory.mb	3776	3776
mapreduce.reduce.memory.mb	7552	7552
yarn.app.mapreduce.am.resource.mb	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Instances m7i

m7i.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7i.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m7i.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

m7i.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

m7i.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	181248	30208
<code>yarn.nodemanager.resource.memory-mb</code>	181248	30208

m7i.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

m7i.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3055m</code>	<code>-Xmx3055m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.map.memory.mb</code>	3819	3819
<code>mapreduce.reduce.memory.mb</code>	7638	7638
<code>yarn.app.mapreduce.am.resource.mb</code>	7638	7638
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30520
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30520

m7i.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx3089m	-Xmx3089m
mapreduce.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.map.memory.mb	3861	3861
mapreduce.reduce.memory.mb	7722	7722
yarn.app.mapreduce.am.resource.mb	7722	7722
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30952
yarn.nodemanager.resource.memory-mb	741376	30952

Instances m7i-flex

m7i-flex.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2342m	-Xmx2342m
mapreduce.java.opts	-Xmx4684m	-Xmx4684m
mapreduce.map.memory.mb	2928	2928
mapreduce.reduce.memory.mb	5856	5856
yarn.app.mapreduce.am.resource.mb	5856	5856

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11712	5856
<code>yarn.nodemanager.resource.memory-mb</code>	11712	5856

m7i-flex.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

m7i-flex.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2714m	-Xmx2714m
mapreduce.java.opts	-Xmx5428m	-Xmx5428m
mapreduce.map.memory.mb	3392	3392
mapreduce.reduce.memory.mb	6784	6784
yarn.app.mapreduce.am.resource.mb	6784	6784
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

m7i-flex.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx2918m	-Xmx2918m
mapreduce.java.opts	-Xmx5836m	-Xmx5836m
mapreduce.map.memory.mb	3648	3648
mapreduce.reduce.memory.mb	7296	7296
yarn.app.mapreduce.am.resource.mb	7296	7296

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

Instances p2

p2.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx21708m</code>	<code>-Xmx21708m</code>
<code>mapreduce.map.memory.mb</code>	13568	13568
<code>mapreduce.reduce.memory.mb</code>	27136	27136
<code>yarn.app.mapreduce.am.resource.mb</code>	27136	27136
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

p2.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24576m</code>	<code>-Xmx24576m</code>
<code>mapreduce.map.memory.mb</code>	15360	15360
<code>mapreduce.reduce.memory.mb</code>	30720	30720
<code>yarn.app.mapreduce.am.resource.mb</code>	30720	30720
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

p2.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx9267m</code>	<code>-Xmx9267m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx18534m</code>	<code>-Xmx18534m</code>
<code>mapreduce.map.memory.mb</code>	11584	11584
<code>mapreduce.reduce.memory.mb</code>	23168	23168
<code>yarn.app.mapreduce.am.resource.mb</code>	23168	23168

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	23168
<code>yarn.nodemanager.resource.memory-mb</code>	741376	23168

Instances p3

p3.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

p3.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

p3.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances p5

p5.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx8294m</code>	<code>-Xmx8294m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx16588m</code>	<code>-Xmx16588m</code>
<code>mapreduce.map.memory.mb</code>	10368	10368
<code>mapreduce.reduce.memory.mb</code>	20736	20736
<code>yarn.app.mapreduce.am.resource.mb</code>	20736	20736
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990656	20736
<code>yarn.nodemanager.resource.memory-mb</code>	1990656	20736

Instances r3

r3.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2342m</code>	<code>-Xmx2342m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx4684m</code>	<code>-Xmx4684m</code>
<code>mapreduce.map.memory.mb</code>	2928	2928
<code>mapreduce.reduce.memory.mb</code>	5856	5856
<code>yarn.app.mapreduce.am.resource.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r3.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2714m</code>	<code>-Xmx2714m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5428m</code>	<code>-Xmx5428m</code>
<code>mapreduce.map.memory.mb</code>	3392	3392
<code>mapreduce.reduce.memory.mb</code>	6784	6784
<code>yarn.app.mapreduce.am.resource.mb</code>	6784	6784

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r3.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx2918m</code>	<code>-Xmx2918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx5836m</code>	<code>-Xmx5836m</code>
<code>mapreduce.map.memory.mb</code>	3648	3648
<code>mapreduce.reduce.memory.mb</code>	7296	7296
<code>yarn.app.mapreduce.am.resource.mb</code>	7296	7296
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r3.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx3021m</code>	<code>-Xmx3021m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.map.memory.mb</code>	3776	3776
<code>mapreduce.reduce.memory.mb</code>	7552	7552
<code>yarn.app.mapreduce.am.resource.mb</code>	7552	7552
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

Instances r4

r4.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r4.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r4.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r4.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r4.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances r5

r5.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4915m</code>	<code>-Xmx4915m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9830m</code>	<code>-Xmx9830m</code>
<code>mapreduce.map.memory.mb</code>	6144	6144
<code>mapreduce.reduce.memory.mb</code>	12288	12288
<code>yarn.app.mapreduce.am.resource.mb</code>	12288	12288
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

r5.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11468m</code>	<code>-Xmx11468m</code>
<code>mapreduce.map.memory.mb</code>	7168	7168
<code>mapreduce.reduce.memory.mb</code>	14336	14336
<code>yarn.app.mapreduce.am.resource.mb</code>	14336	14336

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

r5.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

r5.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6349m</code>	<code>-Xmx6349m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12698m</code>	<code>-Xmx12698m</code>
<code>mapreduce.map.memory.mb</code>	7936	7936
<code>mapreduce.reduce.memory.mb</code>	15872	15872
<code>yarn.app.mapreduce.am.resource.mb</code>	15872	15872
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253 952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253 952	31744

r5.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

r5.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6451m</code>	<code>-Xmx6451m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12902m</code>	<code>-Xmx12902m</code>
<code>mapreduce.map.memory.mb</code>	8064	8064
<code>mapreduce.reduce.memory.mb</code>	16128	16128
<code>yarn.app.mapreduce.am.resource.mb</code>	16128	16128
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	516096	32256
<code>yarn.nodemanager.resource.memory-mb</code>	516096	32256

r5.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6486m</code>	<code>-Xmx6486m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12972m</code>	<code>-Xmx12972m</code>
<code>mapreduce.map.memory.mb</code>	8107	8107
<code>mapreduce.reduce.memory.mb</code>	16214	16214
<code>yarn.app.mapreduce.am.resource.mb</code>	16214	16214
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	778240	32396
<code>yarn.nodemanager.resource.memory-mb</code>	778240	32396

Instances r5a

r5a.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4915m</code>	<code>-Xmx4915m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9830m</code>	<code>-Xmx9830m</code>
<code>mapreduce.map.memory.mb</code>	6144	6144
<code>mapreduce.reduce.memory.mb</code>	12288	12288
<code>yarn.app.mapreduce.am.resource.mb</code>	12288	12288

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

r5a.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11468m</code>	<code>-Xmx11468m</code>
<code>mapreduce.map.memory.mb</code>	7168	7168
<code>mapreduce.reduce.memory.mb</code>	14336	14336
<code>yarn.app.mapreduce.am.resource.mb</code>	14336	14336
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

r5a.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

r5a.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6349m</code>	<code>-Xmx6349m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12698m</code>	<code>-Xmx12698m</code>
<code>mapreduce.map.memory.mb</code>	7936	7936
<code>mapreduce.reduce.memory.mb</code>	15872	15872
<code>yarn.app.mapreduce.am.resource.mb</code>	15872	15872

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253 952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253 952	31744

r5a.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

r5a.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6451m</code>	<code>-Xmx6451m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12902m</code>	<code>-Xmx12902m</code>
<code>mapreduce.map.memory.mb</code>	8064	8064
<code>mapreduce.reduce.memory.mb</code>	16128	16128
<code>yarn.app.mapreduce.am.resource.mb</code>	16128	16128
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	516096	32256
<code>yarn.nodemanager.resource.memory-mb</code>	516096	32256

r5a.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6486m</code>	<code>-Xmx6486m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12972m</code>	<code>-Xmx12972m</code>
<code>mapreduce.map.memory.mb</code>	8107	8107
<code>mapreduce.reduce.memory.mb</code>	16214	16214
<code>yarn.app.mapreduce.am.resource.mb</code>	16214	16214

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	778240	32396
<code>yarn.nodemanager.resource.memory-mb</code>	778240	32396

instances r5ad

r5ad.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r5ad.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r5ad.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r5ad.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r5ad.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r5ad.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6246m	-Xmx6246m
mapreduce.java.opts	-Xmx12492m	-Xmx12492m
mapreduce.map.memory.mb	7808	7808
mapreduce.reduce.memory.mb	15616	15616
yarn.app.mapreduce.am.resource.mb	15616	15616

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	499712	31232
<code>yarn.nodemanager.resource.memory-mb</code>	499712	31232

r5ad.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

instances r5b

r5b.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r5b.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r5b.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r5b.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r5b.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r5b.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r5b.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

Instances r5d

r5d.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx4915m	-Xmx4915m
mapreduce.java.opts	-Xmx9830m	-Xmx9830m
mapreduce.map.memory.mb	6144	6144
mapreduce.reduce.memory.mb	12288	12288
yarn.app.mapreduce.am.resource.mb	12288	12288

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource-memory-mb</code>	24576	12288

r5d.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11468m</code>	<code>-Xmx11468m</code>
<code>mapreduce.map.memory.mb</code>	7168	7168
<code>mapreduce.reduce.memory.mb</code>	14336	14336
<code>yarn.app.mapreduce.am.resource.mb</code>	14336	14336
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource-memory-mb</code>	57344	28672

r5d.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	122880	30720
<code>yarn.nodemanager.resource.memory-mb</code>	122880	30720

r5d.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6349m</code>	<code>-Xmx6349m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12698m</code>	<code>-Xmx12698m</code>
<code>mapreduce.map.memory.mb</code>	7936	7936
<code>mapreduce.reduce.memory.mb</code>	15872	15872
<code>yarn.app.mapreduce.am.resource.mb</code>	15872	15872

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253 952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253 952	31744

r5d.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

r5d.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6451m	-Xmx6451m
mapreduce.java.opts	-Xmx12902m	-Xmx12902m
mapreduce.map.memory.mb	8064	8064
mapreduce.reduce.memory.mb	16128	16128
yarn.app.mapreduce.am.resource.mb	16128	16128
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	516096	32256
yarn.nodemanager.resource.memory-mb	516096	32256

r5d.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6486m	-Xmx6486m
mapreduce.java.opts	-Xmx12972m	-Xmx12972m
mapreduce.map.memory.mb	8107	8107
mapreduce.reduce.memory.mb	16214	16214
yarn.app.mapreduce.am.resource.mb	16214	16214

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	778240	32396
<code>yarn.nodemanager.resource.memory-mb</code>	778240	32396

instances r5dn

r5dn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r5dn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r5dn.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r5dn.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r5dn.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r5dn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r5dn.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

Instances r5n

r5n.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r5n.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r5n.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r5n.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r5n.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r5n.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r5n.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

Instances r6a

r6a.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6a.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6a.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r6a.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6a.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6a.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r6a.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

r6a.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6195m</code>	<code>-Xmx6195m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.map.memory.mb</code>	7744	7744
<code>mapreduce.reduce.memory.mb</code>	15488	15488
<code>yarn.app.mapreduce.am.resource.mb</code>	15488	15488
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

r6a.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	15530	15530
yarn.app.mapreduce.am.resource.mb	15530	15530
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

instances r6g

r6g.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6g.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6g.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r6g.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6g.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6g.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

Instances r6gd

r6gd.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6gd.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6gd.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r6gd.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6gd.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6gd.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances r6i

r6i.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6i.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6i.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r6i.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6i.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6i.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	491520	30720
yarn.nodemanager.resource.memory-mb	491520	30720

r6i.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

r6i.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6029m</code>	<code>-Xmx6029m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12058m</code>	<code>-Xmx12058m</code>
<code>mapreduce.map.memory.mb</code>	7536	7536
<code>mapreduce.reduce.memory.mb</code>	15072	15072
<code>yarn.app.mapreduce.am.resource.mb</code>	15072	15072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	30144
<code>yarn.nodemanager.resource.memory-mb</code>	964608	30144

Instances r6id

r6id.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6id.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6id.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r6id.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r6id.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6id.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r6id.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r6id.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

instances r6idn

r6idn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6idn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r6idn.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r6idn.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6idn.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	366592	30564
yarn.nodemanager.resource.memory-mb	366592	30564

r6idn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6144m	-Xmx6144m
mapreduce.java.opts	-Xmx12288m	-Xmx12288m
mapreduce.map.memory.mb	7680	7680
mapreduce.reduce.memory.mb	15360	15360
yarn.app.mapreduce.am.resource.mb	15360	15360

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r6idn.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

r6idn.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

Instances r6in

r6in.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx4685m	-Xmx4685m
mapreduce.java.opts	-Xmx9370m	-Xmx9370m
mapreduce.map.memory.mb	5856	5856
mapreduce.reduce.memory.mb	11712	11712
yarn.app.mapreduce.am.resource.mb	11712	11712

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r6in.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r6in.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

r6in.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r6in.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r6in.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r6in.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

r6in.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6195m</code>	<code>-Xmx6195m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.map.memory.mb</code>	7744	7744
<code>mapreduce.reduce.memory.mb</code>	15488	15488
<code>yarn.app.mapreduce.am.resource.mb</code>	15488	15488
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

Instances r7a

r7a.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r7a.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r7a.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r7a.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r7a.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r7a.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r7a.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6178m	-Xmx6178m
mapreduce.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.map.memory.mb	7723	7723
mapreduce.reduce.memory.mb	15446	15446
yarn.app.mapreduce.am.resource.mb	15446	15446
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30860
yarn.nodemanager.resource.memory-mb	741376	30860

r7a.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	7744	7744
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

r7a.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6212m</code>	<code>-Xmx6212m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12424m</code>	<code>-Xmx12424m</code>
<code>mapreduce.map.memory.mb</code>	7765	7765
<code>mapreduce.reduce.memory.mb</code>	15530	15530
<code>yarn.app.mapreduce.am.resource.mb</code>	15530	15530
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1490944	31124
<code>yarn.nodemanager.resource.memory-mb</code>	1490944	31124

instances r7g

r7g.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r7g.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r7g.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r7g.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r7g.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r7g.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances r7gd

r7gd.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r7gd.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r7gd.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r7gd.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6042m	-Xmx6042m
mapreduce.java.opts	-Xmx12084m	-Xmx12084m
mapreduce.map.memory.mb	7552	7552
mapreduce.reduce.memory.mb	15104	15104
yarn.app.mapreduce.am.resource.mb	15104	15104
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	30208
yarn.nodemanager.resource.memory-mb	241664	30208

r7gd.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx6110m	-Xmx6110m
mapreduce.java.opts	-Xmx12220m	-Xmx12220m
mapreduce.map.memory.mb	7637	7637
mapreduce.reduce.memory.mb	15274	15274
yarn.app.mapreduce.am.resource.mb	15274	15274

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r7gd.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

Instances R7i

r7i.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r7i.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5427m</code>	<code>-Xmx5427m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.map.memory.mb</code>	6784	6784
<code>mapreduce.reduce.memory.mb</code>	13568	13568
<code>yarn.app.mapreduce.am.resource.mb</code>	13568	13568

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

r7i.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5837m</code>	<code>-Xmx5837m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.map.memory.mb</code>	7296	7296
<code>mapreduce.reduce.memory.mb</code>	14592	14592
<code>yarn.app.mapreduce.am.resource.mb</code>	14592	14592
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r7i.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r7i.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r7i.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r7i.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6178m</code>	<code>-Xmx6178m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12356m</code>	<code>-Xmx12356m</code>
<code>mapreduce.map.memory.mb</code>	7723	7723
<code>mapreduce.reduce.memory.mb</code>	15446	15446
<code>yarn.app.mapreduce.am.resource.mb</code>	15446	15446
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	741376	30860
<code>yarn.nodemanager.resource.memory-mb</code>	741376	30860

r7i.48xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6212m</code>	<code>-Xmx6212m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12424m</code>	<code>-Xmx12424m</code>
<code>mapreduce.map.memory.mb</code>	7765	7765
<code>mapreduce.reduce.memory.mb</code>	15530	15530
<code>yarn.app.mapreduce.am.resource.mb</code>	15530	15530

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1490944	31124
<code>yarn.nodemanager.resource.memory-mb</code>	1490944	31124

Instances r7iz

r7iz.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4685m</code>	<code>-Xmx4685m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9370m</code>	<code>-Xmx9370m</code>
<code>mapreduce.map.memory.mb</code>	5856	5856
<code>mapreduce.reduce.memory.mb</code>	11712	11712
<code>yarn.app.mapreduce.am.resource.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.memory-mb</code>	23424	11712

r7iz.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx5427m	-Xmx5427m
mapreduce.java.opts	-Xmx10854m	-Xmx10854m
mapreduce.map.memory.mb	6784	6784
mapreduce.reduce.memory.mb	13568	13568
yarn.app.mapreduce.am.resource.mb	13568	13568
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	54272	27136
yarn.nodemanager.resource.memory-mb	54272	27136

r7iz.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx5837m	-Xmx5837m
mapreduce.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.map.memory.mb	7296	7296
mapreduce.reduce.memory.mb	14592	14592
yarn.app.mapreduce.am.resource.mb	14592	14592

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

r7iz.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6042m</code>	<code>-Xmx6042m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12084m</code>	<code>-Xmx12084m</code>
<code>mapreduce.map.memory.mb</code>	7552	7552
<code>mapreduce.reduce.memory.mb</code>	15104	15104
<code>yarn.app.mapreduce.am.resource.mb</code>	15104	15104
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

r7iz.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6110m</code>	<code>-Xmx6110m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12220m</code>	<code>-Xmx12220m</code>
<code>mapreduce.map.memory.mb</code>	7637	7637
<code>mapreduce.reduce.memory.mb</code>	15274	15274
<code>yarn.app.mapreduce.am.resource.mb</code>	15274	15274
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	366592	30564
<code>yarn.nodemanager.resource.memory-mb</code>	366592	30564

r7iz.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6144m</code>	<code>-Xmx6144m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.map.memory.mb</code>	7680	7680
<code>mapreduce.reduce.memory.mb</code>	15360	15360
<code>yarn.app.mapreduce.am.resource.mb</code>	15360	15360

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

r7iz.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6029m</code>	<code>-Xmx6029m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12058m</code>	<code>-Xmx12058m</code>
<code>mapreduce.map.memory.mb</code>	7536	7536
<code>mapreduce.reduce.memory.mb</code>	15072	15072
<code>yarn.app.mapreduce.am.resource.mb</code>	15072	15072
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	30144
<code>yarn.nodemanager.resource.memory-mb</code>	964608	30144

Instances x1

x1.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx12058m</code>	<code>-Xmx12058m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24116m</code>	<code>-Xmx24116m</code>
<code>mapreduce.map.memory.mb</code>	15072	15072
<code>mapreduce.reduce.memory.mb</code>	30144	30144
<code>yarn.app.mapreduce.am.resource.mb</code>	30144	30144
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	30144
<code>yarn.nodemanager.resource.memory-mb</code>	964608	30144

x1.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx12109m</code>	<code>-Xmx12109m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24218m</code>	<code>-Xmx24218m</code>
<code>mapreduce.map.memory.mb</code>	15136	15136
<code>mapreduce.reduce.memory.mb</code>	30272	30272
<code>yarn.app.mapreduce.am.resource.mb</code>	30272	30272

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1937408	30272
<code>yarn.nodemanager.resource.memory-mb</code>	1937408	30272

Instances x1e

x1e.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx22682m</code>	<code>-Xmx22682m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx45364m</code>	<code>-Xmx45364m</code>
<code>mapreduce.map.memory.mb</code>	28352	28352
<code>mapreduce.reduce.memory.mb</code>	56704	56704
<code>yarn.app.mapreduce.am.resource.mb</code>	56704	56704
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	113408	0
<code>yarn.nodemanager.resource.memory-mb</code>	113408	0

x1e.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx23501m</code>	<code>-Xmx23501m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx47002m</code>	<code>-Xmx47002m</code>
<code>mapreduce.map.memory.mb</code>	29376	29376
<code>mapreduce.reduce.memory.mb</code>	58752	58752
<code>yarn.app.mapreduce.am.resource.mb</code>	58752	58752
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	235008	0
<code>yarn.nodemanager.resource.memory-mb</code>	235008	0

x1e.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx23910m</code>	<code>-Xmx23910m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx47820m</code>	<code>-Xmx47820m</code>
<code>mapreduce.map.memory.mb</code>	29888	29888
<code>mapreduce.reduce.memory.mb</code>	59776	59776
<code>yarn.app.mapreduce.am.resource.mb</code>	59776	59776

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	478208	0
<code>yarn.nodemanager.resource.memory-mb</code>	478208	0

x1e.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx24115m</code>	<code>-Xmx24115m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx48230m</code>	<code>-Xmx48230m</code>
<code>mapreduce.map.memory.mb</code>	30144	30144
<code>mapreduce.reduce.memory.mb</code>	60288	60288
<code>yarn.app.mapreduce.am.resource.mb</code>	60288	60288
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	0
<code>yarn.nodemanager.resource.memory-mb</code>	964608	0

x1e.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx24218m	-Xmx24218m
mapreduce.java.opts	-Xmx48436m	-Xmx48436m
mapreduce.map.memory.mb	30272	30272
mapreduce.reduce.memory.mb	60544	60544
yarn.app.mapreduce.am.resource.mb	60544	60544
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1937408	0
yarn.nodemanager.resource.memory-mb	1937408	0

x1e.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx24269m	-Xmx24269m
mapreduce.java.opts	-Xmx48538m	-Xmx48538m
mapreduce.map.memory.mb	30336	30336
mapreduce.reduce.memory.mb	60672	60672
yarn.app.mapreduce.am.resource.mb	60672	60672

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	3883008	0
<code>yarn.nodemanager.resource.memory-mb</code>	3883008	0

Instances x2gd

x2gd.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx10854m</code>	<code>-Xmx10854m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx21708m</code>	<code>-Xmx21708m</code>
<code>mapreduce.map.memory.mb</code>	13568	13568
<code>mapreduce.reduce.memory.mb</code>	27136	27136
<code>yarn.app.mapreduce.am.resource.mb</code>	27136	27136
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	54272	27136
<code>yarn.nodemanager.resource.memory-mb</code>	54272	27136

x2gd.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx11674m</code>	<code>-Xmx11674m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx23348m</code>	<code>-Xmx23348m</code>
<code>mapreduce.map.memory.mb</code>	14592	14592
<code>mapreduce.reduce.memory.mb</code>	29184	29184
<code>yarn.app.mapreduce.am.resource.mb</code>	29184	29184
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	29184
<code>yarn.nodemanager.resource.memory-mb</code>	116736	29184

x2gd.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx12083m</code>	<code>-Xmx12083m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24166m</code>	<code>-Xmx24166m</code>
<code>mapreduce.map.memory.mb</code>	15104	15104
<code>mapreduce.reduce.memory.mb</code>	30208	30208
<code>yarn.app.mapreduce.am.resource.mb</code>	30208	30208

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	241664	30208
<code>yarn.nodemanager.resource.memory-mb</code>	241664	30208

x2gd.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx12288m</code>	<code>-Xmx12288m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24576m</code>	<code>-Xmx24576m</code>
<code>mapreduce.map.memory.mb</code>	15360	15360
<code>mapreduce.reduce.memory.mb</code>	30720	30720
<code>yarn.app.mapreduce.am.resource.mb</code>	30720	30720
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	30720
<code>yarn.nodemanager.resource.memory-mb</code>	491520	30720

x2gd.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx12356m	-Xmx12356m
mapreduce.java.opts	-Xmx24712m	-Xmx24712m
mapreduce.map.memory.mb	15445	15445
mapreduce.reduce.memory.mb	30890	30890
yarn.app.mapreduce.am.resource.mb	30890	30890
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	741376	30906
yarn.nodemanager.resource.memory-mb	741376	30906

x2gd.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.java.opts	-Xmx24780m	-Xmx24780m
mapreduce.map.memory.mb	15488	15488
mapreduce.reduce.memory.mb	30976	30976
yarn.app.mapreduce.am.resource.mb	30976	30976

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

instances x2idn

x2idn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx12390m</code>	<code>-Xmx12390m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24780m</code>	<code>-Xmx24780m</code>
<code>mapreduce.map.memory.mb</code>	15488	15488
<code>mapreduce.reduce.memory.mb</code>	30976	30976
<code>yarn.app.mapreduce.am.resource.mb</code>	30976	30976
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	30976
<code>yarn.nodemanager.resource.memory-mb</code>	991232	30976

x2idn.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx12425m</code>	<code>-Xmx12425m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24850m</code>	<code>-Xmx24850m</code>
<code>mapreduce.map.memory.mb</code>	15531	15531
<code>mapreduce.reduce.memory.mb</code>	31062	31062
<code>yarn.app.mapreduce.am.resource.mb</code>	31062	31062
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1490944	31030
<code>yarn.nodemanager.resource.memory-mb</code>	1490944	31030

x2idn.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx12442m</code>	<code>-Xmx12442m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx24884m</code>	<code>-Xmx24884m</code>
<code>mapreduce.map.memory.mb</code>	15552	15552
<code>mapreduce.reduce.memory.mb</code>	31104	31104
<code>yarn.app.mapreduce.am.resource.mb</code>	31104	31104

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990656	31104
<code>yarn.nodemanager.resource.memory-mb</code>	1990656	31104

Instances x2iedn

x2iedn.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx23347m</code>	<code>-Xmx23347m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx46694m</code>	<code>-Xmx46694m</code>
<code>mapreduce.map.memory.mb</code>	29184	29184
<code>mapreduce.reduce.memory.mb</code>	58368	58368
<code>yarn.app.mapreduce.am.resource.mb</code>	58368	58368
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	116736	0
<code>yarn.nodemanager.resource.memory-mb</code>	116736	0

x2iedn.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx24166m	-Xmx24166m
mapreduce.java.opts	-Xmx48332m	-Xmx48332m
mapreduce.map.memory.mb	30208	30208
mapreduce.reduce.memory.mb	60416	60416
yarn.app.mapreduce.am.resource.mb	60416	60416
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	241664	0
yarn.nodemanager.resource.memory-mb	241664	0

x2iedn.4xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
mapreduce.map.java.opts	-Xmx24576m	-Xmx24576m
mapreduce.java.opts	-Xmx49152m	-Xmx49152m
mapreduce.map.memory.mb	30720	30720
mapreduce.reduce.memory.mb	61440	61440
yarn.app.mapreduce.am.resource.mb	61440	61440

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	491520	0
<code>yarn.nodemanager.resource.memory-mb</code>	491520	0

x2iedn.8xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx24781m</code>	<code>-Xmx24781m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49562m</code>	<code>-Xmx49562m</code>
<code>mapreduce.map.memory.mb</code>	30976	30976
<code>mapreduce.reduce.memory.mb</code>	61952	61952
<code>yarn.app.mapreduce.am.resource.mb</code>	61952	61952
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	991232	0
<code>yarn.nodemanager.resource.memory-mb</code>	991232	0

x2iedn.16xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx24883m</code>	<code>-Xmx24883m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49766m</code>	<code>-Xmx49766m</code>
<code>mapreduce.map.memory.mb</code>	31104	31104
<code>mapreduce.reduce.memory.mb</code>	62208	62208
<code>yarn.app.mapreduce.am.resource.mb</code>	62208	62208
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990656	0
<code>yarn.nodemanager.resource.memory-mb</code>	1990656	0

x2iedn.24xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx24918m</code>	<code>-Xmx24918m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49836m</code>	<code>-Xmx49836m</code>
<code>mapreduce.map.memory.mb</code>	31147	31147
<code>mapreduce.reduce.memory.mb</code>	62294	62294
<code>yarn.app.mapreduce.am.resource.mb</code>	62294	62294

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	2990080	-32
<code>yarn.nodemanager.resource.memory-mb</code>	2990080	-32

x2iedn.32xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx24934m</code>	<code>-Xmx24934m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx49868m</code>	<code>-Xmx49868m</code>
<code>mapreduce.map.memory.mb</code>	31168	31168
<code>mapreduce.reduce.memory.mb</code>	62336	62336
<code>yarn.app.mapreduce.am.resource.mb</code>	62336	62336
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	3989504	0
<code>yarn.nodemanager.resource.memory-mb</code>	3989504	0

instances z1d

z1d.xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx4915m</code>	<code>-Xmx4915m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx9830m</code>	<code>-Xmx9830m</code>
<code>mapreduce.map.memory.mb</code>	6144	6144
<code>mapreduce.reduce.memory.mb</code>	12288	12288
<code>yarn.app.mapreduce.am.resource.mb</code>	12288	12288
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	24576	12288
<code>yarn.nodemanager.resource.memory-mb</code>	24576	12288

z1d.2xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx5734m</code>	<code>-Xmx5734m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx11468m</code>	<code>-Xmx11468m</code>
<code>mapreduce.map.memory.mb</code>	7168	7168
<code>mapreduce.reduce.memory.mb</code>	14336	14336
<code>yarn.app.mapreduce.am.resource.mb</code>	14336	14336

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	57344	28672
<code>yarn.nodemanager.resource.memory-mb</code>	57344	28672

z1d.3xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6007m</code>	<code>-Xmx6007m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12014m</code>	<code>-Xmx12014m</code>
<code>mapreduce.map.memory.mb</code>	7509	7509
<code>mapreduce.reduce.memory.mb</code>	15018	15018
<code>yarn.app.mapreduce.am.resource.mb</code>	15018	15018
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	90112	30040
<code>yarn.nodemanager.resource.memory-mb</code>	90112	30040

z1d.6xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6281m</code>	<code>-Xmx6281m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12562m</code>	<code>-Xmx12562m</code>
<code>mapreduce.map.memory.mb</code>	7851	7851
<code>mapreduce.reduce.memory.mb</code>	15702	15702
<code>yarn.app.mapreduce.am.resource.mb</code>	15702	15702
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	188416	31396
<code>yarn.nodemanager.resource.memory-mb</code>	188416	31396

z1d.12xlarge

Option de configuration	Valeur par défaut	Avec HBase installé
<code>mapreduce.map.java.opts</code>	<code>-Xmx6417m</code>	<code>-Xmx6417m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx12834m</code>	<code>-Xmx12834m</code>
<code>mapreduce.map.memory.mb</code>	8021	8021
<code>mapreduce.reduce.memory.mb</code>	16042	16042
<code>yarn.app.mapreduce.am.resource.mb</code>	16042	16042

Option de configuration	Valeur par défaut	Avec HBase installé
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	385024	32100
<code>yarn.nodemanager.resource.memory-mb</code>	385024	32100

Paramètres de configuration de démon Hadoop

Les paramètres de démon Hadoop varient en fonction du type d'instance EC2 utilisé par un nœud de cluster. Les tableaux suivants répertorient les paramètres de configuration par défaut pour chaque type d'instance EC2.

Pour personnaliser ces paramètres, utilisez la classification de configuration `hadoop-env`. Pour plus d'informations, voir la rubrique [Configuration des applications](#).

Types d'instances

- [Instances c1](#)
- [Instances c3](#)
- [Instances c4](#)
- [Instances c5](#)
- [Instances c5a](#)
- [Instances c5ad](#)
- [Instances c5d](#)
- [Instances c5n](#)
- [Instances c6a](#)
- [instances c6g](#)
- [instances c6gd](#)
- [Instances c6gn](#)
- [Instances c6i](#)

- [Instances c6id](#)
- [Instances c6in](#)
- [instances C7a](#)
- [instances c7g](#)
- [Instances c7gd](#)
- [Instances c7gn](#)
- [Instances c7i](#)
- [Instances d2](#)
- [Instances d3](#)
- [instances d3en](#)
- [Instances g3](#)
- [Instances g3s](#)
- [Instances g4dn](#)
- [Instances g5](#)
- [Instances h1](#)
- [Instances i2](#)
- [Instances i3](#)
- [Instances i3en](#)
- [Instances i4g](#)
- [Instances i4i](#)
- [Instances im4gn](#)
- [Instances is4gen](#)
- [Instances m1](#)
- [Instances m2](#)
- [Instances m3](#)
- [Instances m4](#)
- [Instances m5](#)
- [Instances m5a](#)
- [instances m5ad](#)

- [Instances m5d](#)
- [instances m5dn](#)
- [instances m5n](#)
- [instances m5zn](#)
- [Instances m6a](#)
- [Instances m6g](#)
- [instances m6gd](#)
- [Instances m6i](#)
- [Instances m6id](#)
- [instances m6idn](#)
- [instances m6in](#)
- [Instances m7a](#)
- [Instances m7g](#)
- [Instances m7gd](#)
- [Instances m7i](#)
- [Instances m7i-flex](#)
- [Instances p2](#)
- [Instances p3](#)
- [Instances p5](#)
- [Instances r3](#)
- [Instances r4](#)
- [Instances r5](#)
- [Instances r5a](#)
- [instances r5ad](#)
- [instances r5b](#)
- [Instances r5d](#)
- [instances r5dn](#)
- [Instances r5n](#)
- [Instances r6a](#)
- [instances r6g](#)

- [Instances r6gd](#)
- [Instances r6i](#)
- [Instances r6id](#)
- [instances r6idn](#)
- [Instances r6in](#)
- [Instances r7a](#)
- [instances r7g](#)
- [Instances r7gd](#)
- [Instances r7i](#)
- [Instances r7iz](#)
- [Instances x1](#)
- [Instances x1e](#)
- [Instances x2gd](#)
- [instances x2idn](#)
- [Instances x2iedn](#)
- [instances z1d](#)

Instances c1

c1.medium

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	192
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	128
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192

Paramètre	Valeur
HADOOP_DATANODE_HEAPSIZE	96

c1.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	768
YARN_PROXYSERVER_HEAPSIZE	384
YARN_NODEMANAGER_HEAPSIZE	512
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	512
HADOOP_NAMENODE_HEAPSIZE	768
HADOOP_DATANODE_HEAPSIZE	384

Instances c3

c3.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972

Paramètre	Valeur
HADOOP_DATANODE_HEAPSIZE	588

c3.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

c3.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

c3.8.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3317
YARN_PROXYSERVER_HEAPSIZE	3317
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3317
HADOOP_NAMENODE_HEAPSIZE	6348
HADOOP_DATANODE_HEAPSIZE	1679

Instances c4

c4.large

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	1152
YARN_PROXYSERVER_HEAPSIZE	1152
YARN_NODEMANAGER_HEAPSIZE	1152
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1152
HADOOP_NAMENODE_HEAPSIZE	576
HADOOP_DATANODE_HEAPSIZE	384

c4.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c4.2.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

c4.4.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

c4.8.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3317
YARN_PROXYSERVER_HEAPSIZE	3317
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3317
HADOOP_NAMENODE_HEAPSIZE	6348
HADOOP_DATANODE_HEAPSIZE	1679

Instances c5

c5.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

c5.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

c5.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

c5.9xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3563
YARN_PROXYSERVER_HEAPSIZE	3563
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3563
HADOOP_NAMENODE_HEAPSIZE	7577
HADOOP_DATANODE_HEAPSIZE	1925

c5.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

c5.18xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5038
YARN_PROXYSERVER_HEAPSIZE	5038
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5038
HADOOP_NAMENODE_HEAPSIZE	14950
HADOOP_DATANODE_HEAPSIZE	3399

c5.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

Instances c5a

c5a.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c5a.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c5a.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c5a.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c5a.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

c5a.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c5a.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

Instances c5ad

c5ad.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c5ad.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c5ad.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c5ad.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c5ad.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c5ad.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c5ad.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

Instances c5d

c5d.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

c5d.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

c5d.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

c5d.9xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3563
YARN_PROXYSERVER_HEAPSIZE	3563
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3563
HADOOP_NAMENODE_HEAPSIZE	7577
HADOOP_DATANODE_HEAPSIZE	1925

c5d.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

c5d.18xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5038
YARN_PROXYSERVER_HEAPSIZE	5038
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5038
HADOOP_NAMENODE_HEAPSIZE	14950
HADOOP_DATANODE_HEAPSIZE	3399

c5d.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

Instances c5n

c5n.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2304
YARN_PROXYSERVER_HEAPSIZE	2304
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2304
HADOOP_NAMENODE_HEAPSIZE	1280
HADOOP_DATANODE_HEAPSIZE	665

c5n.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2519
YARN_PROXYSERVER_HEAPSIZE	2519
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2519
HADOOP_NAMENODE_HEAPSIZE	2355
HADOOP_DATANODE_HEAPSIZE	880

c5n.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2949
YARN_PROXYSERVER_HEAPSIZE	2949
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2949
HADOOP_NAMENODE_HEAPSIZE	4505
HADOOP_DATANODE_HEAPSIZE	1310

c5n.9xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

c5n.18xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

Instances c6a

c6a.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6a.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6a.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6a.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6a.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6a.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6a.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c6a.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

c6a.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

instances c6g

c6g.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6g.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6g.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6g.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6g.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6g.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

instances c6gd

c6gd.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6gd.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6gd.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6gd.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6gd.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6gd.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

Instances c6gn

c6gn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6gn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6gn.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6gn.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6gn.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6gn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

Instances c6i

c6i.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6i.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6i.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6i.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6i.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6i.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6i.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c6i.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

Instances c6id

c6id.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6id.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6id.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6id.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6id.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6id.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6id.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c6id.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

Instances c6in

c6in.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c6in.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c6in.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c6in.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c6in.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c6in.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c6in.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c6in.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

instances C7a

c7a.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7a.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7a.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7a.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7a.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7a.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c7a.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c7a.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

c7a.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

instances c7g

c7g.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7g.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7g.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7g.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7g.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7g.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

Instances c7gd

c7gd.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7gd.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7gd.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7gd.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7gd.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7gd.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

Instances c7gn

c7gn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7gn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7gn.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7gn.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7gn.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7gn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

Instances c7i

c7i.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972
HADOOP_DATANODE_HEAPSIZE	588

c7i.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

c7i.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

c7i.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

c7i.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

c7i.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

c7i.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

c7i.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

Instances d2

d2.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

d2.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

d2.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

d2.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

Instances d3

d3.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

d3.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

d3.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

d3.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

instances d3en

d3en.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

d3en.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

d3en.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

d3en.6xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

d3en.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

d3en.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

Instances g3

g3.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

g3.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

g3.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances g3s

g3s.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

Instances g4dn

g4dn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

g4dn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

g4dn.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

g4dn.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

g4dn.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

g4dn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

Instances g5

g5.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

g5.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

g5.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

g5.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

g5.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

g5.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

g5.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

g5.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

Instances h1

h1.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

h1.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

h1.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

h1.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

Instances i2

i2.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

i2.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

i2.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

i2.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

Instances i3

i3.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

i3.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

i3.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

i3.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

i3.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances i3en

i3en.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

i3en.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

i3en.3xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

i3en.6xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

i3en.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

i3en.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

Instances i4g

i4g.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

i4g.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

i4g.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

i4g.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

i4g.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances i4i

i4i.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

i4i.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

i4i.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

i4i.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

i4i.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

i4i.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

i4i.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

i4i.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

Instances im4gn

im4gn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

im4gn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

im4gn.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

im4gn.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

im4gn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

Instances is4gen

is4gen.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2557
YARN_PROXYSERVER_HEAPSIZE	2557
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2557
HADOOP_NAMENODE_HEAPSIZE	2547
HADOOP_DATANODE_HEAPSIZE	919

is4gen.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3025
YARN_PROXYSERVER_HEAPSIZE	3025
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3025
HADOOP_NAMENODE_HEAPSIZE	4889
HADOOP_DATANODE_HEAPSIZE	1387

is4gen.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

is4gen.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

Instances m1

m1.small

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	256
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	192
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192
HADOOP_DATANODE_HEAPSIZE	96

m1.medium

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	384
YARN_PROXYSERVER_HEAPSIZE	192
YARN_NODEMANAGER_HEAPSIZE	256
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	256
HADOOP_NAMENODE_HEAPSIZE	384
HADOOP_DATANODE_HEAPSIZE	192

m1.large

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	768
YARN_PROXYSERVER_HEAPSIZE	384
YARN_NODEMANAGER_HEAPSIZE	512
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	512
HADOOP_NAMENODE_HEAPSIZE	768
HADOOP_DATANODE_HEAPSIZE	384

m1.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	1024
YARN_PROXYSERVER_HEAPSIZE	512
YARN_NODEMANAGER_HEAPSIZE	768
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1024
HADOOP_NAMENODE_HEAPSIZE	2304
HADOOP_DATANODE_HEAPSIZE	384

Instances m2

m2.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	1536
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1024
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1024
HADOOP_NAMENODE_HEAPSIZE	3072
HADOOP_DATANODE_HEAPSIZE	384

m2.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	1536
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1024
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1536
HADOOP_NAMENODE_HEAPSIZE	6144
HADOOP_DATANODE_HEAPSIZE	384

m2.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2048
YARN_PROXYSERVER_HEAPSIZE	1024
YARN_NODEMANAGER_HEAPSIZE	1536
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	1536
HADOOP_NAMENODE_HEAPSIZE	12288
HADOOP_DATANODE_HEAPSIZE	384

Instances m3

m3.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

m3.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2703
YARN_PROXYSERVER_HEAPSIZE	2703
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2703
HADOOP_NAMENODE_HEAPSIZE	3276
HADOOP_DATANODE_HEAPSIZE	1064

Instances m4

m4.large

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2252
YARN_PROXYSERVER_HEAPSIZE	2252
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2252
HADOOP_NAMENODE_HEAPSIZE	1024
HADOOP_DATANODE_HEAPSIZE	614

m4.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

m4.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

m4.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

m4.10xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5365
YARN_PROXYSERVER_HEAPSIZE	5365
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5365
HADOOP_NAMENODE_HEAPSIZE	16588
HADOOP_DATANODE_HEAPSIZE	3727

m4.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

Instances m5

m5.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

m5.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

m5.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

m5.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

m5.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

m5.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

m5.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

Instances m5a

m5a.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

m5a.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

m5a.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

m5a.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

m5a.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

m5a.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

m5a.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

instances m5ad

m5ad.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m5ad.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m5ad.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m5ad.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m5ad.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m5ad.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m5ad.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

Instances m5d

m5d.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2416
YARN_PROXYSERVER_HEAPSIZE	2416
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2416
HADOOP_NAMENODE_HEAPSIZE	1843
HADOOP_DATANODE_HEAPSIZE	778

m5d.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

m5d.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

m5d.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

m5d.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

m5d.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

m5d.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

instances m5dn

m5dn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m5dn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m5dn.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m5dn.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m5dn.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m5dn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m5dn.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

instances m5n

m5n.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m5n.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m5n.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m5n.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m5n.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m5n.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m5n.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

instances m5zn

m5zn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2396
YARN_PROXYSERVER_HEAPSIZE	2396
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2396
HADOOP_NAMENODE_HEAPSIZE	1740
HADOOP_DATANODE_HEAPSIZE	757

m5zn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m5zn.3xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3025
YARN_PROXYSERVER_HEAPSIZE	3025
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3025
HADOOP_NAMENODE_HEAPSIZE	4889
HADOOP_DATANODE_HEAPSIZE	1387

m5zn.6xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3962
YARN_PROXYSERVER_HEAPSIZE	3962
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3962
HADOOP_NAMENODE_HEAPSIZE	9574
HADOOP_DATANODE_HEAPSIZE	2324

m5zn.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

Instances m6a

m6a.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6a.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6a.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6a.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6a.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m6a.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6a.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6a.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

m6a.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

Instances m6g

m6g.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6g.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6g.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6g.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6g.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m6g.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

instances m6gd

m6gd.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6gd.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6gd.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6gd.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6gd.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m6gd.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

Instances m6i

m6i.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6i.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6i.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6i.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6i.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m6i.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6i.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6i.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances m6id

m6id.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6id.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6id.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6id.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6id.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m6id.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6id.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6id.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

instances m6idn

m6idn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6idn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6idn.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6idn.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6idn.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m6idn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6idn.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6idn.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

instances m6in

m6in.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m6in.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m6in.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m6in.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m6in.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m6in.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m6in.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m6in.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances m7a

m7a.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7a.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7a.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7a.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m7a.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m7a.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m7a.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m7a.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

m7a.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

Instances m7g

m7g.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7g.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7g.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7g.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m7g.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m7g.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

Instances m7gd

m7gd.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7gd.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7gd.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7gd.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m7gd.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5836
YARN_PROXYSERVER_HEAPSIZE	5836
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5836
HADOOP_NAMENODE_HEAPSIZE	18944
HADOOP_DATANODE_HEAPSIZE	4096

m7gd.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

Instances m7i

m7i.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7i.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7i.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7i.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

m7i.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	5877
YARN_PROXYSERVER_HEAPSIZE	5877
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	5877
HADOOP_NAMENODE_HEAPSIZE	19148
HADOOP_DATANODE_HEAPSIZE	4096

m7i.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

m7i.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

m7i.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

Instances m7i-flex

m7i-flex.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2401
YARN_PROXYSERVER_HEAPSIZE	2401
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2401
HADOOP_NAMENODE_HEAPSIZE	1766
HADOOP_DATANODE_HEAPSIZE	762

m7i-flex.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

m7i-flex.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

m7i-flex.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

Instances p2

p2.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

p2.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

p2.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

Instances p3

p3.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

p3.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

p3.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances p5

p5.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

Instances r3

r3.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r3.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r3.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r3.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

Instances r4

r4.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r4.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r4.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r4.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r4.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances r5

r5.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

r5.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

r5.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

r5.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

r5.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

r5.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

r5.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

Instances r5a

r5a.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

r5a.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

r5a.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

r5a.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

r5a.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

r5a.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

r5a.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

instances r5ad

r5ad.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r5ad.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r5ad.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r5ad.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r5ad.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r5ad.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12247
YARN_PROXYSERVER_HEAPSIZE	12247
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12247
HADOOP_NAMENODE_HEAPSIZE	50995
HADOOP_DATANODE_HEAPSIZE	4096

r5ad.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

instances r5b

r5b.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r5b.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r5b.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r5b.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r5b.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r5b.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r5b.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

Instances r5d

r5d.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

r5d.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

r5d.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4710
YARN_PROXYSERVER_HEAPSIZE	4710
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4710
HADOOP_NAMENODE_HEAPSIZE	13312
HADOOP_DATANODE_HEAPSIZE	3072

r5d.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7331
YARN_PROXYSERVER_HEAPSIZE	7331
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7331
HADOOP_NAMENODE_HEAPSIZE	26419
HADOOP_DATANODE_HEAPSIZE	4096

r5d.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

r5d.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12574
YARN_PROXYSERVER_HEAPSIZE	12574
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12574
HADOOP_NAMENODE_HEAPSIZE	52633
HADOOP_DATANODE_HEAPSIZE	4096

r5d.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17817
YARN_PROXYSERVER_HEAPSIZE	17817
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17817
HADOOP_NAMENODE_HEAPSIZE	78848
HADOOP_DATANODE_HEAPSIZE	4096

instances r5dn

r5dn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r5dn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r5dn.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r5dn.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r5dn.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r5dn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r5dn.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

Instances r5n

r5n.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r5n.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r5n.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r5n.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r5n.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r5n.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r5n.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

Instances r6a

r6a.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6a.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6a.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6a.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6a.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6a.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6a.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6a.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

r6a.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

instances r6g

r6g.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6g.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6g.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6g.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6g.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6g.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances r6gd

r6gd.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6gd.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6gd.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6gd.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6gd.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6gd.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances r6i

r6i.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6i.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6i.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6i.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6i.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6i.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6i.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6i.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

Instances r6id

r6id.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6id.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6id.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6id.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6id.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6id.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6id.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6id.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

instances r6idn

r6idn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6idn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6idn.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6idn.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6idn.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6idn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6idn.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6idn.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

Instances r6in

r6in.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r6in.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r6in.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r6in.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r6in.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r6in.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r6in.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r6in.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

Instances r7a

r7a.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7a.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7a.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7a.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7a.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7a.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r7a.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r7a.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

r7a.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

instances r7g

r7g.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7g.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7g.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7g.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7g.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7g.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances r7gd

r7gd.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7gd.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7gd.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7gd.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7gd.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7gd.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

Instances r7i

r7i.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7i.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7i.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7i.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7i.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7i.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r7i.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

r7i.48xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

Instances r7iz

r7iz.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2713
YARN_PROXYSERVER_HEAPSIZE	2713
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2713
HADOOP_NAMENODE_HEAPSIZE	3328
HADOOP_DATANODE_HEAPSIZE	1075

r7iz.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

r7iz.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

r7iz.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

r7iz.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9584
YARN_PROXYSERVER_HEAPSIZE	9584
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9584
HADOOP_NAMENODE_HEAPSIZE	37683
HADOOP_DATANODE_HEAPSIZE	4096

r7iz.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

r7iz.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

Instances x1

x1.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

x1.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	41000
YARN_PROXYSERVER_HEAPSIZE	41000
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	41000
HADOOP_NAMENODE_HEAPSIZE	194764
HADOOP_DATANODE_HEAPSIZE	4096

Instances x1e

x1e.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4520
YARN_PROXYSERVER_HEAPSIZE	4520
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4520
HADOOP_NAMENODE_HEAPSIZE	12364
HADOOP_DATANODE_HEAPSIZE	2882

x1e.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6952
YARN_PROXYSERVER_HEAPSIZE	6952
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6952
HADOOP_NAMENODE_HEAPSIZE	24524
HADOOP_DATANODE_HEAPSIZE	4096

x1e.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	11816
YARN_PROXYSERVER_HEAPSIZE	11816
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	11816
HADOOP_NAMENODE_HEAPSIZE	48844
HADOOP_DATANODE_HEAPSIZE	4096

x1e.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	21544
YARN_PROXYSERVER_HEAPSIZE	21544
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	21544
HADOOP_NAMENODE_HEAPSIZE	97484
HADOOP_DATANODE_HEAPSIZE	4096

x1e.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	41000
YARN_PROXYSERVER_HEAPSIZE	41000
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	41000
HADOOP_NAMENODE_HEAPSIZE	194764
HADOOP_DATANODE_HEAPSIZE	4096

x1e.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	79912
YARN_PROXYSERVER_HEAPSIZE	79912
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	79912
HADOOP_NAMENODE_HEAPSIZE	389324
HADOOP_DATANODE_HEAPSIZE	4096

Instances x2gd

x2gd.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3338
YARN_PROXYSERVER_HEAPSIZE	3338
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3338
HADOOP_NAMENODE_HEAPSIZE	6451
HADOOP_DATANODE_HEAPSIZE	1699

x2gd.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

x2gd.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

x2gd.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

x2gd.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	17080
YARN_PROXYSERVER_HEAPSIZE	17080
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	17080
HADOOP_NAMENODE_HEAPSIZE	75161
HADOOP_DATANODE_HEAPSIZE	4096

x2gd.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

instances x2idn

x2idn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

x2idn.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	32071
YARN_PROXYSERVER_HEAPSIZE	32071
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	32071
HADOOP_NAMENODE_HEAPSIZE	150118
HADOOP_DATANODE_HEAPSIZE	4096

x2idn.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

Instances x2iedn

x2iedn.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4587
YARN_PROXYSERVER_HEAPSIZE	4587
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4587
HADOOP_NAMENODE_HEAPSIZE	12697
HADOOP_DATANODE_HEAPSIZE	2949

x2iedn.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	7086
YARN_PROXYSERVER_HEAPSIZE	7086
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	7086
HADOOP_NAMENODE_HEAPSIZE	25190
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.4xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	12083
YARN_PROXYSERVER_HEAPSIZE	12083
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	12083
HADOOP_NAMENODE_HEAPSIZE	50176
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.8xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	22077
YARN_PROXYSERVER_HEAPSIZE	22077
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	22077
HADOOP_NAMENODE_HEAPSIZE	100147
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.16xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	42065
YARN_PROXYSERVER_HEAPSIZE	42065
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	42065
HADOOP_NAMENODE_HEAPSIZE	200089
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.24xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	62054
YARN_PROXYSERVER_HEAPSIZE	62054
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	62054
HADOOP_NAMENODE_HEAPSIZE	300032
HADOOP_DATANODE_HEAPSIZE	4096

x2iedn.32xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	82042
YARN_PROXYSERVER_HEAPSIZE	82042
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	82042
HADOOP_NAMENODE_HEAPSIZE	399974
HADOOP_DATANODE_HEAPSIZE	4096

instances z1d

z1d.xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	2744
YARN_PROXYSERVER_HEAPSIZE	2744
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2744
HADOOP_NAMENODE_HEAPSIZE	3481
HADOOP_DATANODE_HEAPSIZE	1105

z1d.2xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	3399
YARN_PROXYSERVER_HEAPSIZE	3399
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	3399
HADOOP_NAMENODE_HEAPSIZE	6758
HADOOP_DATANODE_HEAPSIZE	1761

z1d.3xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	4055
YARN_PROXYSERVER_HEAPSIZE	4055
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	4055
HADOOP_NAMENODE_HEAPSIZE	10035
HADOOP_DATANODE_HEAPSIZE	2416

z1d.6xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	6021
YARN_PROXYSERVER_HEAPSIZE	6021
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	6021
HADOOP_NAMENODE_HEAPSIZE	19865
HADOOP_DATANODE_HEAPSIZE	4096

z1d.12xlarge

Paramètre	Valeur
YARN_RESOURCEMANAGER_HEAPSIZE	9953
YARN_PROXYSERVER_HEAPSIZE	9953
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	9953
HADOOP_NAMENODE_HEAPSIZE	39526
HADOOP_DATANODE_HEAPSIZE	4096

HDFSconfiguration

Le tableau suivant décrit les paramètres par défaut du système de fichiers distribué Hadoop (HDFS) et leurs paramètres. Vous pouvez modifier ces valeurs à l'aide de la classification de configuration `hdfs-site`. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

Warning

1. La `dfs.replication` valeur 1 sur les clusters de moins de quatre nœuds peut entraîner une perte de HDFS données en cas de panne d'un seul nœud. Si votre cluster dispose d'un HDFS espace de stockage, nous vous recommandons de le configurer avec au moins quatre nœuds principaux pour les charges de travail de production afin d'éviter toute perte de données.
2. Amazon n'EMR autorisera pas les clusters à dimensionner les nœuds principaux situés en dessous `dfs.replication`. Par exemple, si `dfs.replication = 2`, le nombre minimum de nœuds principaux est 2.
3. Lorsque vous utilisez la mise à l'échelle gérée, autoscaling, ou que vous choisissiez de redimensionner manuellement votre cluster, nous vous recommandons de définir `dfs.replication` sur une valeur supérieure ou égale à 2.

Paramètre	Définition	Valeur par défaut
<code>dfs.block.size</code>	La taille des HDFS blocs. Lorsque vous travaillez sur des données stockées dans HDFS, la taille divisée est généralement la taille d'un HDFS bloc. Plus la taille est grande, plus la granularité de la tâche est faible. En revanche, les contraintes sur le NameNode du cluster sont moins importantes.	134217728 (128 Mo)
<code>dfs.replication</code>	Le nombre de copies de chaque bloc à stocker pour des questions de durabilité. Amazon EMR définit cette valeur en fonction du nombre de nœuds principaux avec lesquels le cluster est approvisionné. Ajustez la valeur en fonction de vos besoins. Pour remplacer la valeur par défaut, utilisez la classification <code>hdfs-site</code> .	1 pour les clusters provisionnés avec moins de quatre nœuds principaux 2 pour les clusters approvisionnés avec moins de dix nœuds principaux 3 pour tous les autres clusters

Chiffrement transparent HDFS sur Amazon EMR

Le chiffrement transparent est mis en œuvre grâce à l'utilisation de zones de HDFS chiffrement, qui sont HDFS des chemins que vous définissez. Chaque zone de chiffrement a sa propre clé, qui est stockée dans le serveur de clés spécifié à l'aide de la classification de configuration `hdfs-site`.

À partir de la EMR version 4.8.0 d'Amazon, vous pouvez utiliser les configurations de EMR sécurité Amazon pour configurer plus facilement les paramètres de chiffrement des données pour les clusters. Les configurations de sécurité proposent des paramètres permettant de sécuriser les données en transit et les données inactives dans les volumes de stockage Amazon Elastic Block Store EBS (Amazon) et les EMRFS données dans Amazon S3. Pour plus d'informations, consultez la section [Chiffrer les données en transit et au repos](#) dans le guide EMR de gestion Amazon.

Amazon EMR utilise Hadoop KMS par défaut ; toutefois, vous pouvez en utiliser un autre KMS qui implémente l' `KeyProvider` API opération. Chaque fichier d'une zone de HDFS chiffrement possède

sa propre clé de chiffrement des données unique, qui est chiffrée par la clé de zone de chiffrement. HDFS Les données sont cryptées end-to-end (au repos et en transit) lorsque les données sont écrites dans une zone de chiffrement, car les activités de chiffrement et de déchiffrement ne se produisent que dans le client.

Vous ne pouvez pas déplacer des fichiers entre les zones de chiffrement ou d'une zone de chiffrement vers des chemins d'accès non chiffrés.

Le HDFS client NameNode et interagit avec le Hadoop KMS (ou un autre KMS que vous avez configuré) au cours de l'opération KeyProvider API. KMS est chargé de stocker les clés de chiffrement dans le keystore de sauvegarde. Amazon EMR inclut également la politique de force JCE illimitée, ce qui vous permet de créer des clés à la longueur souhaitée.

Pour plus d'informations, consultez la section [Chiffrement transparent HDFS dans](#) la documentation Hadoop.

Note

Dans Amazon EMR, la fonction KMS over HTTPS n'est pas activée par défaut avec Hadoop KMS. Pour plus d'informations sur la façon d'activer le KMS over HTTPS, consultez la documentation [Hadoop. KMS](#)

Configuration du chiffrement HDFS transparent

Vous pouvez configurer le chiffrement transparent dans Amazon EMR en créant des clés et en ajoutant des zones de chiffrement. Pour ce faire, plusieurs options s'offrent à vous :

- Utilisation de l'opération EMR de configuration Amazon lors de la création d'un cluster
- Utiliser une JAR étape Hadoop avec `command-runner.jar`
- En vous connectant au nœud principal du cluster Hadoop et en utilisant les clients de ligne de commande `hadoop key` et `hdfs crypto`
- En utilisant le REST APIs pour Hadoop KMS et HDFS

Pour plus d'informations sur le REST APIs, consultez la documentation correspondante pour Hadoop KMS et HDFS

Pour créer des zones de chiffrement et leurs clés lors de la création du cluster à l'aide du CLI

La `hdfs-encryption-zones` classification dans l'API opération de configuration vous permet de spécifier un nom de clé et une zone de chiffrement lorsque vous créez un cluster. Amazon EMR crée cette clé dans Hadoop KMS sur votre cluster et configure la zone de chiffrement.

- Créez un cluster à l'aide de la commande suivante.

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

myConfig.json:

```
[
  {
    "Classification": "hdfs-encryption-zones",
    "Properties": {
      "/myHDFSPath1": "path1_key",
      "/myHDFSPath2": "path2_key"
    }
  }
]
```

Pour créer des zones de chiffrement et leurs clés manuellement sur le nœud maître

1. Lancez votre cluster à l'aide d'une EMR version d'Amazon supérieure à 4.1.0.
2. Connectez-vous au nœud principal du cluster avec SSH.
3. Créez une clé dans HadoopKMS.

```
$ hadoop key create path2_key
```

```
path2_key has been successfully created with options Options{cipher='AES/CTR/
NoPadding', bitLength=256, description='null', attributes=null}.
KMSClientProvider[http://ip-x-x-x-x.ec2.internal:16000/kms/v1/] has been updated.
```

Important

Hadoop KMS exige que les noms de vos clés soient en minuscules. Si vous utilisez une clé comprenant des majuscules, votre cluster échouera lors du lancement.

4. Créez le chemin de la zone de chiffrement dans HDFS.

```
$ hadoop fs -mkdir /myHDFSPath2
```

5. Faites du HDFS chemin une zone de chiffrement à l'aide de la clé que vous avez créée.

```
$ hdfs crypto -createZone -keyName path2_key -path /myHDFSPath2
Added encryption zone /myHDFSPath2
```

Pour créer des zones de chiffrement et leurs clés manuellement à l'aide du AWS CLI

- Ajoutez des étapes pour créer les KMS clés et les zones de chiffrement manuellement à l'aide de la commande suivante.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps Type=CUSTOM_JAR,Name="Create
  First Hadoop KMS Key",Jar="command-runner.jar",ActionOnFailure=CONTINUE,Args=[/
bin/bash,-c,"\"hadoop key create path1_key\""] \
Type=CUSTOM_JAR,Name="Create First Hadoop HDFS Path",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop fs -mkdir /
myHDFSPath1\""] \
Type=CUSTOM_JAR,Name="Create First Encryption Zone",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hdfs crypto -createZone
-keyName path1_key -path /myHDFSPath1\""] \
Type=CUSTOM_JAR,Name="Create Second Hadoop KMS Key",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop key create
path2_key\""] \
Type=CUSTOM_JAR,Name="Create Second Hadoop HDFS Path",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hadoop fs -mkdir /
myHDFSPath2\""] \
```



```
Type=CUSTOM_JAR,Name="Create Second Encryption Zone",Jar="command-runner.jar",ActionOnFailure=CONTINUE,Args=[/bin/bash,-c,"\"hdfs crypto -createZone -keyName path2_key -path /myHDFSPath2\""]
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

Considérations relatives au chiffrement HDFS transparent

Une bonne pratique consiste à créer une zone de chiffrement pour chaque application où des fichiers peuvent être écrits. Vous pouvez également tout chiffrer HDFS en utilisant la `hdfs-encryption-zones` classification dans la configuration API et en spécifiant le chemin racine (/) comme zone de chiffrement.

Serveur de gestion des clés Hadoop

[Hadoop KMS](#) est un serveur de gestion de clés qui permet d'implémenter des services cryptographiques pour les clusters Hadoop et qui peut servir de fournisseur clé pour [Chiffrement transparent HDFS sur Amazon EMR](#). Hadoop dans KMS Amazon EMR est installé et activé par défaut lorsque vous sélectionnez l'application Hadoop lors du lancement d'un cluster. EMR Hadoop KMS ne stocke pas les clés lui-même, sauf dans le cas d'une mise en cache temporaire. Hadoop KMS agit comme un proxy entre le fournisseur de clés et le mandataire du client vers un magasin de clés de sauvegarde ; il ne s'agit pas d'un magasin de clés. Le keystore par défaut créé pour Hadoop KMS est l'extension de cryptographie Java (). KeyStore JCEKS La politique de résistance JCE illimitée est également incluse, ce qui vous permet de créer des clés de la longueur souhaitée. Hadoop prend KMS également en charge une gamme de solutions ACLs qui contrôlent l'accès aux clés et aux opérations clés indépendamment des autres applications clientes, telles que. HDFS La longueur de clé par défaut dans Amazon EMR est de 256 bits.

Pour configurer HadoopKMS, utilisez la `hadoop-kms-site` classification pour modifier les paramètres. Pour configurerACLs, vous utilisez la classification `kms-acls`.

Pour plus d'informations, consultez la documentation [Hadoop. KMS](#). Hadoop KMS est utilisé dans le chiffrement transparent HDFS Hadoop. Pour en savoir plus sur le chiffrement HDFS transparent,

consultez la rubrique relative au [chiffrement HDFS transparent](#) dans la documentation d'Apache Hadoop.

Note

Dans AmazonEMR, la fonction KMS over n'HTTPS est pas activée par défaut avec HadoopKMS. Pour savoir comment activer le KMS overHTTPS, consultez la documentation [Hadoop. KMS](#)

Important

Hadoop KMS exige que les noms de vos clés soient en minuscules. Si vous utilisez une clé comprenant des majuscules, votre cluster échouera lors du lancement.

Configuration de Hadoop KMS dans Amazon EMR

Si vous utilisez la EMR version 4.6.0 ou ultérieure d'Amazon, il `kms-http-port` s'agit de la version 9700 et `kms-admin-port` de la version 9701.

Vous pouvez configurer Hadoop KMS au moment de la création du cluster à l'aide de la configuration des API versions AmazonEMR. Les classifications d'objets de configuration disponibles pour KMS Hadoop sont les suivantes :

Classifications de configuration Hadoop KMS

Classification	Nom de fichier
hadoop-kms-site	<code>kms-site.xml</code>
hadoop-kms-acls	<code>kms-acls.xml</code>
hadoop-kms-env	<code>kms-env.sh</code>
hadoop-kms-log4j	<code>kms-log4j.properties</code>

Pour configurer Hadoop KMS ACLs à l'aide du CLI

- Créez un cluster avec Hadoop à ACLs l'KMSaide de la commande suivante :

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-acls",
    "Properties": {
      "hadoop.kms.blacklist.CREATE": "hdfs,foo,myBannedUser",
      "hadoop.kms.acl.ROLLOVER": "myAllowedUser"
    }
  }
]
```

Pour désactiver le KMS cache Hadoop à l'aide du CLI

- Créez un cluster avec Hadoop KMS `hadoop.kms.cache.enable` défini sur `false`, à l'aide de la commande suivante :

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.cache.enable": "false"
    }
  }
]
```

Pour définir des variables d'environnement dans le **kms-env.sh** script à l'aide du CLI

- Modifiez les paramètres dans **kms-env.sh** via la configuration **hadoop-kms-env**. Créez un cluster avec Hadoop à KMS l'aide de la commande suivante :

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=App1 Name=App2 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-env",
    "Properties": {
    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "JAVA_LIBRARY_PATH": "/path/to/files",
          "KMS_SSL_KEYSTORE_FILE": "/non/Default/Path/.keystore",
          "KMS_SSL_KEYSTORE_PASS": "myPass"
        },
        "Configurations": [
        ]
      }
    ]
  }
]
```

Pour plus d'informations sur la configuration d'HadoopKMS, consultez la documentation [KMSHadoop](#).

HDFSchiffrement transparent sur les EMR clusters comportant plusieurs nœuds principaux

[Apache Ranger](#) KMS est utilisé dans un EMR cluster Amazon comportant plusieurs nœuds principaux pour un chiffrement transparentHDFS.

Apache Ranger KMS stocke sa clé racine et ses clés de zone de chiffrement (EZ) dans votre Amazon RDS pour un EMR cluster Amazon comportant plusieurs nœuds principaux. Pour activer HDFS le chiffrement transparent dans un EMR cluster Amazon comportant plusieurs nœuds principaux, vous devez fournir les configurations suivantes.

- Amazon RDS ou votre propre connexion My SQL Server URL pour stocker la clé KMS racine Ranger et la clé EZ
- Nom d'utilisateur et mot de passe pour My SQL
- Mot de passe pour la clé KMS root Ranger
- PEMFichier d'autorité de certification (CA) pour SSL la connexion à Mon SQL serveur

Vous pouvez fournir ces configurations à l'aide des classifications `ranger-kms-dbks-site` et `ranger-kms-db-ca`, comme l'illustre l'exemple suivant.

```
[
  {
    "Classification": "ranger-kms-dbks-site",
    "Properties": {
      "ranger.ks.jpa.jdbc.url": "jdbc:log4jdbc:mysql://mysql-host-url.xxx-xxx-1.xxx.amazonaws.com:3306/rangerkms",
      "ranger.ks.jpa.jdbc.user": "mysql-user-name",
      "ranger.ks.jpa.jdbc.password": "mysql-password",
      "ranger.db.encrypt.key.password": "password-for-encrypting-a-master-key"
    }
  },
  {
    "Classification": "ranger-kms-db-ca",
    "Properties": {
      "ranger.kms.trust.ca.file.s3.url": "s3://rds-downloads/rds-ca-2019-root.pem"
    }
  }
]
```

Voici les classifications des objets de configuration pour Apache RangerKMS.

Classifications de configuration Hadoop KMS

Classification	Description
<code>ranger-kms-dbks-site</code>	Modifiez les valeurs dans le fichier <code>dbks-site.xml</code> de RangerKMS.
<code>ranger-kms-site</code>	Modifiez les valeurs dans le fichier <code>ranger-kms-site.xml</code> de RangerKMS.
<code>ranger-kms-env</code>	Modifiez les valeurs dans l'KMSenvironnement Ranger.
<code>ranger-kms-log4j</code>	Modifiez les valeurs dans le fichier <code>kms-log4j.properties</code> de Ranger. KMS
<code>ranger-kms-db-ca</code>	Modifiez les valeurs du fichier CA sur S3 pour Ma SQL SSL connexion avec RangerKMS.

Considérations

- Il est vivement recommandé de chiffrer votre RDS instance Amazon pour améliorer la sécurité. Pour plus d'informations, consultez [Présentation du chiffrement des RDS ressources Amazon](#).
- Il est vivement recommandé d'utiliser une base de SQL données My Database distincte pour chaque EMR cluster Amazon avec plusieurs nœuds principaux pour une barre de sécurité élevée.
- Pour configurer HDFS le chiffrement transparent dans un EMR cluster Amazon comportant plusieurs nœuds principaux, vous devez spécifier la `hdfs-encryption-zones` classification lors de la création du cluster. Dans le cas contraire, Ranger ne KMS sera ni configuré ni démarré. La reconfiguration de la `hdfs-encryption-zones` classification ou de l'une des classifications de KMS configuration Hadoop sur un cluster en cours d'exécution n'est pas prise en charge sur un EMR cluster Amazon comportant plusieurs nœuds principaux.

Création ou exécution d'une application Hadoop

Rubriques

- [Créer des fichiers binaires à l'aide d'Amazon EMR](#)
- [Traitement des données avec Streaming](#)
- [Traitez les données avec un outil personnalisé JAR](#)

Créer des fichiers binaires à l'aide d'Amazon EMR

Vous pouvez utiliser Amazon EMR comme un environnement de génération pour compiler les programmes à utiliser dans votre cluster. Les programmes que vous utilisez avec Amazon EMR doivent être compilés sur un système exécutant la même version de Linux que celle utilisée par Amazon EMR. Pour une version 32 bits, les options de compilation sur un ordinateur 32 bits ou de compilation croisée 32 bits doivent être activées. Pour une version 64 bits, les options de compilation sur un ordinateur 64 bits ou de compilation croisée 64 bits doivent être activées. Pour plus de plus amples informations sur les versions d'instance EC2, consultez [Planification et configuration des instances](#) dans le Guide de gestion d'Amazon EMR. Les langages de programmation pris en charge incluent C++, Python et C#.

Le tableau suivant décrit les étapes nécessaires pour générer et tester votre application à l'aide d'Amazon EMR.

Processus de création d'un module

- 1 Connectez-vous au nœud maître de votre cluster.
- 2 Copiez les fichiers source dans le nœud principal.
- 3 Créez des fichiers binaires avec les optimisations nécessaires.
- 4 Copiez des fichiers binaires à partir du nœud principal dans Amazon S3.

Chacune de ces étapes est détaillée dans les sections suivantes.

Pour vous connecter au nœud principal du cluster

- Suivez les instructions fournies dans [Connexion au nœud principal à l'aide de SSH](#) dans le Guide de gestion d'Amazon EMR.

Pour copier les fichiers source dans le nœud principal

1. Placez vos fichiers source dans un compartiment Amazon S3. Pour savoir comment créer des compartiments et comment transférer des données vers Amazon S3, consultez le [Guide du l'utilisateur Amazon Simple Storage Service](#).
2. Créez un dossier sur votre cluster Hadoop pour vos fichiers source en entrant une commande similaire à celle qui suit :

```
mkdir SourceFiles
```

3. Copiez vos fichiers source à partir d'Amazon S3 vers le nœud principal en tapant une commande similaire à celle qui suit :

```
hadoop fs -get s3://mybucket/SourceFiles SourceFiles
```

Pour créer des fichiers binaires avec les optimisations nécessaires

La façon dont vous créez vos fichiers binaires dépend de plusieurs facteurs. Suivez les instructions correspondant aux outils d'installation que vous utilisez pour installer et configurer votre environnement. Vous pouvez utiliser les commandes de spécification du système Hadoop

pour obtenir les informations relatives au cluster, afin de déterminer comment installer votre environnement.

Pour identifier les spécifications du système

- Utilisez les commandes suivantes pour vérifier l'architecture que vous utilisez pour créer vos fichiers binaires.
 - a. Pour afficher la version de Debian, saisissez la commande suivante :

```
master$ cat /etc/issue
```

La sortie ressemble à ce qui suit.

```
Debian GNU/Linux 5.0
```

- b. Pour afficher le nom DNS public et la taille du processeur, saisissez la commande suivante :

```
master$ uname -a
```

La sortie ressemble à ce qui suit.

```
Linux domU-12-31-39-17-29-39.compute-1.internal 2.6.21.7-2.fc8xen #1 SMP Fri  
Feb 15 12:34:28 EST 2008 x86_64 GNU/Linux
```

- c. Pour afficher la vitesse du processeur, saisissez la commande suivante :

```
master$ cat /proc/cpuinfo
```

La sortie ressemble à ce qui suit.

```
processor : 0  
vendor_id : GenuineIntel  
model name : Intel(R) Xeon(R) CPU E5430 @ 2.66GHz  
flags : fpu tsc msr pae mce cx8 apic mca cmov pat pse36 clflush dts acpi mmx  
fxsr sse sse2 ss ht tm syscall nx lm constant_tsc pni monitor ds_cpl vmx est  
tm2 ssse3 cx16 xtpr cda lahf_lm  
...
```

Une fois que vos fichiers binaires sont créés, vous pouvez copier les fichiers dans Amazon S3.

Pour copier des fichiers binaires à partir du nœud principal dans Amazon S3

- Entrez la commande suivante pour copier les fichiers binaires dans votre compartiment Amazon S3 :

```
hadoop fs -put BinaryFiles s3://mybucket/BinaryDestination
```

Traitement des données avec Streaming

Le streaming Hadoop est un utilitaire fourni avec Hadoop qui vous permet de développer des MapReduce exécutables dans des langages autres que Java. Le streaming est implémenté sous la forme d'un JAR fichier, vous pouvez donc l'exécuter depuis Amazon EMR API ou la ligne de commande comme un JAR fichier standard.

Cette section explique comment utiliser le streaming avec AmazonEMR.

Note

Le streaming Apache Hadoop est un outil indépendant. En tant que tel, ses fonctions et ses paramètres ne sont pas tous décrits ici. Pour plus d'informations sur le streaming Hadoop, rendez-vous sur <http://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming.html>.

Utilisation de l'utilitaire de streaming Hadoop

Cette section décrit comment utiliser l'utilitaire de streaming Hadoop.

Processus Hadoop

- 1 Ecrivez votre exécutable mappeur et réducteur dans le langage de programmation de votre choix.

Suivez les instructions dans la documentation Hadoop pour écrire vos exécutable de diffusion en continu. Les programmes doivent lire leur entrée à partir de données d'entrée et de sortie standard par le biais de sortie standard. Par défaut, chaque ligne d'entrée/

sortie représente un enregistrement et le premier onglet sur chaque ligne est utilisé comme séparateur entre la clé et la valeur.

- 2 Testez vos exécutables localement et chargez-les sur Amazon S3.
- 3 Utilisez l'interface de ligne de commande Amazon EMR ou la console Amazon EMR pour exécuter votre application.

Chaque script mappeur se lance comme un processus distinct dans le cluster. Chaque exécutable de réducteur remplace le résultat de l'exécutable du mappeur par la sortie de données du flux de travail.

Les paramètres `input`, `output`, `mapper`, et `reducer` sont requis par la plupart des applications de streaming. Le tableau suivant décrit ces paramètres et d'autres, facultatifs.

Paramètre	Description	Obligatoire
-input	<p>Emplacement des données d'entrée sur Amazon S3.</p> <p>Type : chaîne</p> <p>Par défaut : aucun</p> <p>Contrainte :URI. Si aucun protocole n'est spécifié, il utilise alors le système de fichiers par défaut du cluster.</p>	Oui
-output	<p>Emplacement sur Amazon S3 où Amazon EMR télécharge les données traitées.</p> <p>Type : chaîne</p> <p>Par défaut : aucun</p> <p>Contrainte : URI</p> <p>Par défaut : si aucun emplacement n'est spécifié, Amazon EMR télécharge les données vers l'emplacement spécifié par <code>input</code>.</p>	Oui
-mapper	Nom de l'exécutable du mappeur.	Oui

Paramètre	Description	Obligatoire
	Type : chaîne Par défaut : aucun	
-reducer	Nom de l'exécutable du réducteur. Type : chaîne Par défaut : aucun	Oui
-cacheFile	Un emplacement Amazon S3 qui contient des fichiers que Hadoop peut copier dans votre répertoire de travail local (principalement pour améliorer les performances). Type : chaîne Par défaut : aucun Contraintes : [URI] # [nom du lien symbolique à créer dans le répertoire de travail]	Non
-cacheArchive	JARfichier à extraire dans le répertoire de travail Type : chaîne Par défaut : aucun Contraintes : [URI] # [nom du répertoire de liens symboliques à créer dans le répertoire de travail]	Non
-combiner	Combine les résultats Type : chaîne Par défaut : aucun Contraintes : nom de classe Java	Non

L'exemple de code suivant est un exécutable de mappeur écrit dans Python. Ce script fait partie de l'WordCount exemple d'application.

```
#!/usr/bin/python
import sys

def main(argv):
    line = sys.stdin.readline()
    try:
        while line:
            line = line.rstrip()
            words = line.split()
            for word in words:
                print "LongValueSum:" + word + "\t" + "1"
            line = sys.stdin.readline()
    except "end of file":
        return None
if __name__ == "__main__":
    main(sys.argv)
```

Soumission d'une étape de streaming

Cette section présente les concepts de base de la soumission d'une étape de streaming à un cluster. Une application de streaming lit l'entrée depuis l'entrée standard, puis exécute un script ou un exécutable (appelé un mappeur) sur chaque entrée. Le résultat de chacune des entrées est enregistré localement, généralement sur une partition Hadoop Distributed File System (HDFS). Une fois que toutes les entrées sont traitées par le mappeur, un second script ou exécutable (appelé un réducteur) traite les résultats du mappeur. Les résultats du réducteur sont envoyés vers la sortie standard. Vous pouvez enchaîner une série d'étapes de streaming, où le résultat d'une étape devient l'entrée d'une autre.

Le mappeur et le réducteur peuvent être chacun référencés en tant que fichier ou vous pouvez fournir une classe Java. Vous pouvez implémenter le mappeur et le réducteur dans tous les langages pris en charge, notamment Ruby, Perl/PHP, Python ou Bash.

Soumission d'une étape de streaming à l'aide de la console

Cet exemple décrit comment utiliser la EMR console Amazon pour soumettre une étape de streaming à un cluster en cours d'exécution.

Pour soumettre une étape de streaming

1. Ouvrez la EMR console Amazon à l'[adresse https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr).
2. Dans la Liste de clusters, sélectionnez le nom de votre cluster.
3. Faites défiler l'affichage jusqu'à la section Étapes et développez-la, puis choisissez Ajouter une étape.
4. Dans la boîte de dialogue Ajouter une étape :
 - Pour Type d'étape, choisissez Programme de streaming.
 - Pour Nom, acceptez le nom par défaut (Programme de streaming) ou saisissez un nouveau nom.
 - Pour Mappeur, saisissez l'emplacement de votre classe de mappeur, ou accédez à ce dernier via le navigateur, dans Hadoop ou dans un compartiment S3 dans lequel l'exécutable de mappeur, tel qu'un programme Python, réside. La valeur du chemin doit être au format *BucketName/path/MapperExecutable*.
 - Pour Reducer (Réducteur), saisissez l'emplacement de votre classe de mappeur, ou accédez à ce dernier via le navigateur, dans Hadoop, ou dans un compartiment S3 dans lequel l'exécutable du réducteur, tel qu'un programme Python, réside. La valeur du chemin doit être au format *BucketName/path/MapperExecutable*. Amazon EMR prend en charge le mot clé agrégé spécial. Pour plus d'informations, consultez la bibliothèque Aggregate fournie par Hadoop.
 - Pour Emplacement S3 d'entrée, saisissez l'emplacement de vos données d'entrée ou accédez à ce dernier via le navigateur.
 - Pour Emplacement S3 de sortie, saisissez le nom du compartiment de sortie Amazon S3 ou naviguez à ce dernier.
 - Pour Arguments, laissez le champ vide.
 - Pour Action sur échec, acceptez l'option par défaut (Continuer).
5. Choisissez Ajouter. L'étape s'affiche dans la console avec le statut En suspens.
6. Le statut de l'étape passe de Pending (En suspens) à Running (En cours d'exécution) puis à Completed (Terminé), au fur et à mesure de son exécution. Pour mettre à jour l'état, choisissez l'icône d'actualisation au-dessus de la colonne Actions.

AWS CLI

Ces exemples montrent comment utiliser l'étape AWS CLI pour créer un cluster et soumettre une étape de streaming.

Pour créer un cluster et soumettre une étape de streaming à l'aide du AWS CLI

- Pour créer un cluster et soumettre une étape de streaming à l'aide du AWS CLI, tapez la commande suivante et remplacez *myKey* avec le nom de votre paire de EC2 clés. Notez que votre argument pour `--files` doit être le chemin Amazon S3 vers l'emplacement de votre script, et les arguments pour `-mapper` et `-reducer` doivent être les noms des fichiers de script respectifs.

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 --
applications Name=Hue Name=Hive Name=Pig --use-default-roles \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=STREAMING,Name="Streaming Program",ActionOnFailure=CONTINUE,Args=[--
files,pathtoscripts,-mapper,mapperscript,-reducer,reducerscript,aggregate,-
input,pathtoinputdata,-output,pathtooutputbucket]
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

Lorsque vous spécifiez le nombre d'instances sans utiliser le paramètre `--instance-groups`, un seul nœud maître est lancé et les instances restantes sont lancées en tant que nœuds principaux. Tous les nœuds utilisent le type d'instance spécifié dans la commande.

Note

Si vous n'avez pas encore créé le rôle de EMR service Amazon et le profil d'EC2instance par défaut, tapez `aws emr create-default-roles` pour les créer avant de saisir la `create-cluster` sous-commande.

Pour plus d'informations sur l'utilisation EMR des commandes Amazon dans le AWS CLI, consultez <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Traitez les données avec un outil personnalisé JAR

Un programme personnalisé JAR exécute un programme Java compilé que vous pouvez télécharger sur Amazon S3. Vous devez compiler le programme en fonction de la version de Hadoop que vous souhaitez lancer et soumettre une CUSTOM_JAR étape à votre cluster AmazonEMR. Pour plus d'informations sur la compilation d'un JAR fichier, consultez [Créer des fichiers binaires à l'aide d'Amazon EMR](#).

Pour plus d'informations sur la création d'une MapReduce application Hadoop, consultez le [MapReduce didacticiel](#) de la documentation d'Apache Hadoop.

Rubriques

- [Soumettre une JAR étape personnalisée](#)

Soumettre une JAR étape personnalisée

Un programme personnalisé JAR exécute un programme Java compilé que vous pouvez télécharger sur Amazon S3. Vous devez compiler le programme en fonction de la version de Hadoop que vous souhaitez lancer et soumettre une CUSTOM_JAR étape à votre cluster AmazonEMR. Pour plus d'informations sur la compilation d'un JAR fichier, consultez [Créer des fichiers binaires à l'aide d'Amazon EMR](#).

Pour plus d'informations sur la création d'une MapReduce application Hadoop, consultez le [MapReduce didacticiel](#) de la documentation d'Apache Hadoop.

Cette section décrit les principes de base de la soumission d'une JAR étape personnalisée sur AmazonEMR. La soumission d'une JAR étape personnalisée vous permet d'écrire un script pour traiter vos données avec le langage de programmation Java.

Soumettre une JAR étape personnalisée avec la console

Cet exemple décrit comment utiliser la EMR console Amazon pour soumettre une JAR étape personnalisée à un cluster en cours d'exécution.

Pour soumettre une JAR étape personnalisée avec la console

1. Ouvrez la EMR console Amazon à l'[adresse https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr).
2. Dans la Liste de clusters, sélectionnez le nom de votre cluster.
3. Faites défiler l'affichage jusqu'à la section Étapes et développez-la, puis choisissez Ajouter une étape.
4. Dans la boîte de dialogue Ajouter une étape :
 - Pour Type d'étape, choisissez Personnalisé JAR.
 - Pour Nom, acceptez le nom par défaut (personnaliséJAR) ou saisissez un nouveau nom.
 - Pour l'emplacement JAR S3, tapez ou naviguez jusqu'à l'emplacement de votre JAR fichier. JARemplacement peut-être un chemin vers S3 ou une classe Java entièrement qualifiée dans le classpath.
 - Pour Arguments, tapez les arguments requis sous forme de chaînes séparées par des espaces ou laissez le champ vide.
 - Pour Action sur échec, acceptez l'option par défaut (Continuer).
5. Choisissez Ajouter. L'étape s'affiche dans la console avec le statut En suspens.
6. Le statut de l'étape passe de Pending (En suspens) à Running (En cours d'exécution) puis à Completed (Terminé), au fur et à mesure de son exécution. Pour mettre à jour l'état, choisissez l'icône d'actualisation au-dessus de la colonne Actions.

Lancer un cluster et soumettre une JAR étape personnalisée à l'aide du AWS CLI

Pour lancer un cluster et soumettre une JAR étape personnalisée à l'aide du AWS CLI

Pour lancer un cluster et soumettre une JAR étape personnalisée avec le AWS CLI, tapez la `create-cluster` sous-commande avec le `--steps` paramètre.

- Pour lancer un cluster et soumettre une JAR étape personnalisée, tapez la commande suivante, remplacez *myKey* avec le nom de votre paire de EC2 clés, et remplacez *mybucket* avec le nom de votre compartiment.

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \  
--applications Name=Hue Name=Hive Name=Pig --use-default-roles \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \  

```

```
--steps Type=CUSTOM_JAR,Name="Custom JAR
Step",ActionOnFailure=CONTINUE,Jar=pathtojarfile,Args=["pathtoinputdata","pathtooutputbucket"]
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

Lorsque vous spécifiez le nombre d'instances sans le paramètre `--instance-groups`, un seul nœud primaire est lancé et les instances restantes sont lancées en tant que nœuds principaux. Tous les nœuds utilisent le type d'instance que vous spécifiez dans la commande.

Note

Si vous n'avez pas encore créé le rôle de EMR service Amazon et le profil d'EC2instance par défaut, tapez `aws emr create-default-roles` pour les créer avant de saisir la `create-cluster` sous-commande.

Pour plus d'informations sur l'utilisation EMR des commandes Amazon dans le AWS CLI, consultez <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Dépendances tierces

Parfois, il peut être nécessaire de l'inclure dans le MapReduce chemin de classe JARs à utiliser avec votre programme. Vous avez deux options pour cela :

- Incluez `--libjars s3://URI_to_JAR` dans les options de l'étape pour la procédure dans [Lancer un cluster et soumettre une JAR étape personnalisée à l'aide du AWS CLI](#).
- Lancez le cluster avec un paramètre `mapreduce.application.classpath` modifié dans `mapred-site.xml`. Utilisez la classification de configuration `mapred-site`. Pour créer le cluster à l'aide de l'étape AWS CLI, cela ressemblerait à ce qui suit :

```
aws emr create-cluster --release-label emr-7.2.0 \
--applications Name=Hue Name=Hive Name=Pig --use-default-roles \
--instance-type m5.xlarge --instance-count 2 --ec2-attributes KeyName=myKey \
```

```
--steps Type=CUSTOM_JAR,Name="Custom JAR
Step",ActionOnFailure=CONTINUE,Jar=pathtojarfile,Args=["pathtoinputdata","pathtooutputbucket
\
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

myConfig.json:

```
[
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapreduce.application.classpath": "path1,path2"
    }
  }
]
```

La liste des chemins séparés par des virgules doit être ajoutée au chemin de JVM classe pour chaque tâche.

Lire les objets restaurés

Avec les EMR versions 7.2.0 et supérieures d'Amazon, vous pouvez lire les objets Glacier restaurés depuis l'emplacement S3 de la table contenant le S3A protocole. Les moteurs des versions précédentes ne font pas de distinction entre les fichiers Glacier et Glacier Deep Archive, ce qui signifie que vous en obteniez un `AmazonS3Exception` si vous essayiez d'accéder à un fichier Glacier en cours avec S3A. Cette opération de lecture ignore les fichiers Glacier archivés s'ils sont toujours en cours de restauration. Pour activer ce comportement, utilisez le paramètre `fs.s3a.glacier.read.restored.objects`. Ce paramètre peut prendre trois valeurs :

- `READ_ALL` — cette valeur indique qu'Amazon EMR ne doit pas prendre en compte les classes de stockage extraites d'Amazon S3. Il s'agit du comportement de par défaut.
- `SKIP_ALL_GLACIER` — cette valeur indique qu'Amazon EMR doit ignorer tous les objets S3 marqués avec la classe de stockage Glacier et récupérer tous les autres objets. Il s'agit du comportement par défaut d'Amazon Athena en ce qui concerne les objets Glacier.
- `READ_RESTORED_GLACIER_OBJECTS` — cette valeur indique qu'Amazon EMR doit vérifier l'état restauré de l'objet Glacier. Si Amazon EMR peut restaurer l'objet, vous pouvez le lire comme un objet S3 normal. Dans le cas contraire, Amazon EMR ignore l'objet de S3 Glacier

Exemples

Spark

Pour lire les objets restaurés lorsque vous utilisez Spark, utilisez la configuration suivante :

```
--conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

Si vous utilisez spark-sql, utilisez plutôt la configuration suivante ;

```
spark-sql --conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

Flink

Si vous utilisez Flink, vous pouvez définir la configuration dans le fichier `flink-conf.yaml`

```
fs.s3a.glacier.read.restored.objects: <value>
```

Vous pouvez également définir le `flink-conf` classement :

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "fs.s3a.glacier.read.restored.objects": "<value>"
    }
  }
]
```

Hive

Si vous utilisez Hive, définissez la configuration dans le `hive-site.xml` fichier.

```
<property>
  <name>fs.s3a.glacier.read.restored.objects</name>
  <value><value></value>
</property>
```

Vous pouvez également utiliser la Hive CLI pour définir la propriété `--hiveconf` :

```
hive --hiveconf fs.s3a.glacier.read.restored.objects=<value>
```

Considérations

Lorsque vous lisez des objets restaurés depuis S3 Glacier, tenez compte des considérations suivantes :

- Vous ne pouvez lire les objets restaurés que si vous utilisez le S3A schéma ou le S3AFileSystem pour accéder aux données.
- Lorsque vous lisez un objet Glacier restauré, Amazon EMR ne le restaure pas lui-même. Pour ce faire, vous devez utiliser le AWS CLI ou le AWS SDK.

Activer la détection non uniforme de l'accès à la mémoire pour les conteneurs YARN

Avec EMR les versions 6.x et ultérieures d'Amazon, vous pouvez utiliser un accès mémoire non uniforme (NUMA) pour le multitraitement de vos données sur des clusters. NUMA est un modèle de conception de la mémoire d'un ordinateur dans lequel le processeur peut accéder à sa propre mémoire locale plus rapidement que la mémoire d'un autre processeur ou partagée entre processeurs. YARN les conteneurs ont de meilleures performances NUMA car ils peuvent se lier à un NUMA nœud spécifique qui gère toutes les allocations de mémoire ultérieures. Cela réduit le nombre de fois que votre cluster doit accéder à la mémoire distante.

Vous pouvez activer la NUMA prise en charge du YARN conteneur lorsque la machine du nœud de travail est une machine à NUMA nœuds multiples. Pour vérifier si un composant master est un nœud NUMA unique ou multi-NUMA, exécutez la commande suivante.

```
lscpu | grep -i numa  
NUMA node(s): 2
```

En général, les instances dont la taille est supérieure à 12 fois possèdent deux nœuds NUMA. Cela ne s'applique pas aux instances en métal.

Pour activer la NUMA sensibilisation aux YARN conteneurs

1. Utilisez la `yarn-site` configuration suivante dans votre cluster Amazon EMR 6.x.

```
[  
  {  
    "classification": "yarn-site",
```

```

        "properties":{
            "yarn.nodemanager.linux-container-executor.nonsecure-mode.local-
user":"yarn",
            "yarn.nodemanager.linux-container-executor.group":"yarn",
            "yarn.nodemanager.container-
executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor",
            "yarn.nodemanager.numa-awareness.enabled":"true",
            "yarn.nodemanager.numa-awareness.numactl.cmd":"/usr/bin/numactl",
            "yarn.nodemanager.numa-awareness.read-topology":"true"
        },
        "configurations":[]
    }
]

```

2. Fournissez l'action d'amorçage suivante dans votre cluster.

```

#!/bin/bash

sudo yum -y install numactl
echo 1 | sudo tee /proc/sys/kernel/numa_balancing

echo "banned.users=mapred,bin,hdfs" >> /etc/hadoop/conf/container-executor.cfg
rm -rf /var/log/hadoop-yarn/
sudo chown -R yarn:hadoop /var/log/hadoop-yarn/
sudo chmod 755 -R /var/log/hadoop-yarn/

sudo chmod 6050 /etc/hadoop/conf/container-executor.cfg

mkdir /mnt/yarn && sudo chmod 755 -R /mnt/yarn && sudo chown -R yarn:hadoop /mnt/
yarn
mkdir /mnt1/yarn && sudo chmod 755 -R /mnt1/yarn && sudo chown -R yarn:hadoop /
mnt1/yarn
mkdir /mnt2/yarn && sudo chmod 755 -R /mnt2/yarn && sudo chown -R yarn:hadoop /
mnt2/yarn

```

3. Chaque conteneur doit être conscient de l'existence de NUMA. Vous pouvez informer la machine virtuelle Java (JVM) de chaque conteneur à l'aide d'un NUMA drapeau. Par exemple, pour indiquer le JVM à utiliser NUMA dans une MapReduce tâche, ajoutez les propriétés suivantes dans `mapred-site.xml`.

```

<property>
  <name>mapreduce.reduce.java.opts</name>
  <value>-XX:+UseNUMA</value>

```

```
</property>
<property>
  <name>mapreduce.map.java.opts</name>
  <value>-XX:+UseNUMA</value>
</property>
```

4. Pour vérifier que vous avez activé NUMA, effectuez une recherche dans l'un des fichiers journaux de NodeManager à l'aide de la commande suivante.

```
grep "NUMA resources allocation is enabled," *
```

Pour vérifier que des ressources de NUMA nœud NodeManager ont été attribuées à un conteneur, effectuez une recherche dans le NodeManager journal à l'aide de la commande suivante, en `<container_id>` remplaçant par votre propre ID de conteneur.

```
grep "NUMA node" | grep <container_id>
```

Historique des versions de Hadoop

Le tableau suivant répertorie la version de Hadoop incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Hadoop

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-7.2.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodema

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.36.2	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-7.1.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-7.0.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.15.0	3.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.14.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.13.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.12.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.11.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.11.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.10.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.10.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.9.1	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.9.0	3.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.8.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.8.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.7.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.36.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.36.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.6.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.35.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.5.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.4.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.3.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.3.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.2.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.2.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.1.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.1.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-6.0.1	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-6.0.0	3.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.34.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.33.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.33.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.32.1	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.32.0	2.10.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.31.1	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.31.0	2.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.30.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.30.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
mr-5.29.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.28.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.28.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.27.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.27.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.26.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.25.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.24.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.24.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.23.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.23.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.22.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.21.2	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.21.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.21.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.20.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.20.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.19.1	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.19.0	2.8.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.18.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.18.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.17.2	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.17.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.17.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.16.1	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.16.0	2.8.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.15.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.15.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.14.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.14.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.14.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.13.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.13.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.3	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.12.2	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.12.1	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.12.0	2.8.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.11.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.11.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.11.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.10.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.10.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.9.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.8.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.8.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.7.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.7.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.6.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.6.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server
emr-5.5.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.5.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.5.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.5.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.4.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.4.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.3.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.3.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.2.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.2.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.2.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.1.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.1.0	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.0.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-5.0.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-5.0.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.6	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-4.9.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.9.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-4.9.1	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.5	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.4	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-4.8.3	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.2	2.7.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.8.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-4.8.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.4	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.3	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-4.7.2	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.7.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-4.6.1	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.6.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.5.0	2.7.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-4.4.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.3.0	2.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.2.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Étiquette Amazon EMR Release	Version de Hadoop	Composants installés avec Hadoop
emr-4.1.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager
emr-4.0.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Notes de mise à jour de Hadoop par version

Rubriques

- [Amazon EMR 7.2.0 - Notes de mise à jour de Hadoop](#)
- [Amazon EMR 6.6.0 - Notes de mise à jour de Hadoop](#)

Amazon EMR 7.2.0 - Notes de mise à jour de Hadoop

Amazon EMR 7.2.0 - Modifications apportées à Hadoop

Type	Description
------	-------------

Type	Description
Nouvelle fonction	HADOOP-18850 : S3A : activer le chiffrement double couche côté serveur avec des clés AWS KMS
Nouvelle fonction	Support S3 Glacier pour lire les objets restaurés
Amélioration	Dimensionnement dynamique du nombre maximal de connexions S3A en fonction des cœurs lors de l'exécution de Spark Jobs
Amélioration	HADOOP-18797 : Support des écritures simultanées avec le S3A Magic Committer
Amélioration	HADOOP-19047 : Support et InMemory suivi des S3A Magic Commits
Amélioration	Ajouter les méthodes « Export », « Publish » et « » AbortPublish dans MagicV2 Committer
Amélioration	Implémenter ReadFullyIntoBuffers FileSystem API dans S3A
Correctif de bogue.	HADOOP-18793 : S3A StagingCommitter ne nettoie pas le répertoire de staging-uploads
Upgrade	HADOOP-19008 : S3A : mise à jour de la version aws-sdk vers la version 2.21.41
Upgrade	HADOOP-18613 : Mise à niveau ZooKeeper vers la version 3.8.3

Amazon EMR 7.2.0 - Fonctionnalités de Hadoop

Consultez la liste suivante pour découvrir les nouvelles fonctionnalités Hadoop d'Amazon EMR 7.2.0.

- [Support de S3 Glacier pour lire les objets restaurés](#) — Avec les EMR versions 7.2.0 et supérieures d'Amazon, vous pouvez lire les objets Glacier restaurés depuis l'emplacement S3 de la table contenant le S3A protocole. Cette opération de lecture ignore les fichiers Glacier archivés s'ils sont toujours en cours de restauration.
- Support des écritures simultanées avec le commutateur magique S3A — Amazon EMR 7.2.0 introduit la prise en charge des écritures simultanées avec le validateur magique S3A. [Cette fonctionnalité a été ajoutée à la communauté OSS Hadoop dans le cadre de HADOOP -18797.](#)

Amazon EMR 6.6.0 - Notes de mise à jour de Hadoop

Amazon EMR 6.6.0 - Modifications apportées à Hadoop

Type	Description
Bogue	Correction d'enregistrements dupliqués lors de la lecture de fichiers BZip2 texte.
Rétroportage	HADOOP-18136 : Vérification de la <code>FileUtils.unTar()</code> gestion des fichiers .tar manquants
Rétroportage	HADOOP-17627 : Rétroportage vers la branche 3.2 HADOOP -17371, -17621, -17625 pour mettre à jour Jetty vers la version HADOOP 9.4.39 HADOOP
Rétroportage	HADOOP-17655 : mise à niveau de Jetty vers la version 9.4.40
Rétroportage	HADOOP-17796 : Mise à niveau de la version Jetty vers la version 9.4.43
Rétroportage	HADOOP-17661 : <code>mvn versions:set</code> ne parvient pas à analyser le fichier pom.xml

Type	Description
Rétroportage	HADOOP-17236 : Augmenter snakeyaml à 1,26 pour atténuer -2017-18640 CVE
Rétroportage	HADOOP-16717 : Supprime GenericsUtil isLog 4 jLogger dépendances sur Log4 jLoggerAdapter
Rétroportage	HADOOP-17633 : Passez json-smart à 2.4.2 et à 9.8 en raison de nimbus-jose-jwt CVEs
Rétroportage	HADOOP-17844 : mise à niveau de JSON Smart vers la version 2.4.7
Rétroportage	HADOOP-17972 : Backport HADOOP -17683 (mise à jour de commons-io vers la version 2.8.0) pour la branche 3.2
Rétroportage	HADOOP-16555 : Mise à jour de commons-compress vers la version 1.19
Rétroportage	HADOOP-17370 : Mise à niveau de commons-compress vers la version 1.21
Rétroportage	HADOOP-17096 : Corrige le décalage de la mémoire tampon ZStandardCompressor d'entrée
Rétroportage	HADOOP-17112 : Les espaces blancs ne sont pas autorisés dans les chemins lors de l'enregistrement de fichiers sur s3a via un validateur
Rétroportage	HADOOP-13500 : Synchronisation de l'itération de l'objet des propriétés de configuration
Rétroportage	HDFS-14099 : Descripteur de trame inconnu lors de la décompression de plusieurs images dans ZStandardDecompressor

Type	Description
Rétroportage	HDFS-16410 : Analyse XML non sécurisée dans OfflineEditsXMLLoader
Rétroportage	HDFS-14498 : LeaseManager peut tourner en boucle indéfiniment sur le fichier dont la création a échoué
Rétroportage	HDFS-15290 : activé NPE au démarrage HttpServer NameNode
Rétroportage	HDFS-15293 : Assouplir la condition d'acceptation d'une image lors de la réception d'un point de contrôle
Rétroportage	HDFS-12979 : StandbyNode doit être téléchargé après le point FsImage de contrôle ObserverNode
Rétroportage	YARN-10538 : Ajout de nœuds de remise en service à la liste des nœuds mis à jour renvoyés à l'AM
Rétroportage	YARN-10472 : Rétroportage YARN -10314 (YarnClientrenvoie uniquement NoClassDefFoundError les WebSocketException fichiers jar clients ombrés) vers la branche 3.2
Rétroportage	YARN-9968 : Public Localizer est en cours de fermeture en raison de NodeManager NullPointerException
Rétroportage	YARN-10651 : CapacityScheduler s'est écrasé avec in NPE AbstractYarnScheduler.updateNodeResource()

Type	Description
Rétroportage	YARN-9339 : La métrique des applications en attente est incorrecte après le déplacement de l'application vers une nouvelle file d'attente
Rétroportage	YARN-10438 : Handle null in container Id ClientRMService#getContainerReport()
Rétroportage	YARN-7266 : ATS 1.5 ne démarre pas si RollingLevelDb les fichiers sont corrompus ou manquants
Rétroportage	YARN-9063 : ATS 1.5 ne démarre pas si RollingLevelDb les fichiers sont corrompus ou manquants
Rétroportage	YARN-9848 : Annuler YARN -4946 (RM ne doit pas considérer une application comme COMPLETED lorsque l'agrégation de journaux n'est pas dans un état terminal).

Apache HBase

[HBase](#) est une base de données distribuée open source, non relationnelle, développée dans le cadre du projet Hadoop de l'Apache Software Foundation. HBase s'exécute sur le système de fichiers distribué Hadoop (HDFS) pour fournir des fonctionnalités de base de données non relationnelles à l'écosystème Hadoop. HBase est inclus dans les EMR versions 4.6.0 et ultérieures d'Amazon.

HBase fonctionne parfaitement avec Hadoop, partage son système de fichiers et sert d'entrée et de sortie directes au MapReduce framework et au moteur d'exécution. HBase s'intègre également à Apache Hive, permettant des requêtes SQL similaires sur des HBase tables, des jointures avec des tables basées sur Hive et la prise en charge de la connectivité de base de données Java (JDBC). Pour plus d'informations sur HBase, consultez [Apache HBase](#) et [HBase la documentation](#) sur le site Web d'Apache. Pour un exemple d'utilisation HBase avec Hive, consultez le billet de blog sur le AWS Big Data [Combine No SQL and Massively parallel analytics using Apache HBase and Apache Hive on Amazon](#). EMR

Avec HBase on Amazon EMR, vous pouvez également sauvegarder vos HBase données directement sur Amazon Simple Storage Service (Amazon S3) et les restaurer à partir d'une sauvegarde créée précédemment lors du lancement HBase d'un cluster. Amazon EMR propose des options supplémentaires d'intégration à Amazon S3 pour la persistance des données et la reprise après sinistre.

- HBase sur Amazon S3 - Avec Amazon EMR version 5.2.0 et versions ultérieures, vous pouvez utiliser Amazon S3 pour stocker HBase le répertoire HBase racine et les métadonnées d'un cluster directement sur Amazon S3. Vous pouvez par la suite démarrer un nouveau cluster, en le désignant à l'emplacement du répertoire racine dans Amazon S3. Un seul cluster à la fois peut utiliser l'emplacement dans Amazon S3, à l'exception d'un cluster en lecture et réplication. Pour de plus amples informations, veuillez consulter [HBase sur Amazon S3 \(mode de stockage Amazon S3\)](#).
- HBase read-replicas - Amazon EMR version 5.7.0 et versions ultérieures avec Amazon HBase S3 prend en charge les clusters de lecture-réplicas. Un cluster réplica en lecture offre un accès en lecture seule aux fichiers et métadonnées de stockage d'un cluster principal pour des opérations en lecture seule. Pour de plus amples informations, veuillez consulter [Utilisation d'un cluster réplica en lecture](#).
- HBase Instantanés - Comme alternative à Amazon S3, avec les EMR versions 4.0 et ultérieures, vous pouvez créer des instantanés de vos HBase données directement HBase sur Amazon S3,

puis récupérer les données à l'aide des instantanés. Pour de plus amples informations, veuillez consulter [Utilisation de HBase snapshots](#).

Important

Pour le dimensionnement des EMR HBase clusters Amazon, nous vous déconseillons d'utiliser le [dimensionnement géré](#) ou le [dimensionnement avec des politiques personnalisées](#) pour les HBase clusters.

Le tableau suivant répertorie la version HBase incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants avec lesquels Amazon EMR installe HBase

Pour la version des composants installés HBase dans cette version, voir [Versions des composants de la version 7.2.0](#).

Informations de version pour emr-7.2.0

Étiquette EMR de sortie Amazon	HBaseVersion	Composants installés avec HBase
emr-7.2.0	HBase 2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Le tableau suivant répertorie la version HBase incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants avec lesquels Amazon EMR installe HBase

Pour connaître la version des composants installés HBase dans cette version, consultez la [version 6.15.0 Versions des composants](#).

HBase informations de version pour emr-6.15.0

Étiquette EMR de sortie Amazon	HBaseVersion	Composants installés avec HBase
emr-6.15.0	HBase 2.4.17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Note

Apache HBase HBCK2 est un outil opérationnel distinct pour réparer HBase les régions et les tables système. Dans les EMR versions 6.1.0 et ultérieures d'Amazon, le fichier `hbase-hbck2.jar` est fourni `/usr/lib/hbase-operator-tools/` sur le nœud principal. Pour plus d'informations sur la création et l'utilisation de l'outil, consultez [HBaseHBCK2](#).

Le tableau suivant répertorie la version HBase incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants avec lesquels Amazon EMR installe. HBase

Pour la version des composants installés HBase dans cette version, voir la [version 5.36.2 Versions des composants](#).

HBase informations de version pour emr-5.36.2

Étiquette EMR de sortie Amazon	HBaseVersion	Composants installés avec HBase
emr-5.36.2	HBase 1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Rubriques

- [Création d'un cluster avec HBase](#)
- [HBasesur Amazon S3 \(mode de stockage Amazon S3\)](#)
- [Journaux d'écriture anticipée \(WAL\) pour Amazon EMR](#)
- [Utilisation de la HBase coque](#)
- [Accédez aux HBase tables avec Hive](#)
- [Utilisation de HBase snapshots](#)
- [Configurez HBase](#)
- [Afficher l'interface HBase utilisateur](#)

- [Afficher les fichiers journaux HBase](#)
- [Moniteur HBase avec Ganglia](#)
- [Migration depuis les versions précédentes HBase](#)
- [HBase historique des versions](#)

Création d'un cluster avec HBase

Les procédures décrites dans cette section décrivent les principes de base du lancement d'un cluster à l'aide du AWS Management Console et du AWS CLI. Pour obtenir des informations détaillées sur la planification, la configuration et le lancement des EMR clusters Amazon, consultez la section [Planification et configuration des clusters](#) dans le guide EMR de gestion Amazon.

Création d'un cluster à HBase l'aide de la console

Pour connaître les étapes rapides à suivre pour lancer des clusters avec la console, consultez [Getting started with Amazon EMR](#) dans le guide EMR de gestion Amazon.

Pour lancer un cluster HBase installé à l'aide de la console

1. Ouvrez la EMR console Amazon à l'adresse <https://console.aws.amazon.com/emr>.
2. Choisissez Create a cluster (Créer un cluster) et Go to advanced options (Accéder aux options avancées).
3. Pour Configuration des logiciels, choisissez une Version d'Amazon 4.6.0 ou ultérieure (nous recommandons la dernière version). Choisissez HBase et d'autres applications comme vous le souhaitez.
4. Avec Amazon EMR version 5.2.0 et versions ultérieures, sous Paramètres HBase de stockage, sélectionnez HDFSS3. Pour de plus amples informations, veuillez consulter [HBasesur Amazon S3 \(mode de stockage Amazon S3\)](#).
5. Sélectionnez d'autres options si nécessaire, puis choisissez Create cluster (Créer le cluster).

Création d'un cluster à HBase l'aide du AWS CLI

Utilisez la commande suivante pour créer un cluster avec HBase installé :

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \  
--applications Name=HBase --use-default-roles --ec2-attributes KeyName=myKey \  

```

```
--instance-type m5.xlarge --instance-count 3
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

Si vous l'utilisez HBase sur Amazon S3, spécifiez l'`--configurationoption` en faisant référence à un objet JSON de configuration. L'objet de configuration doit contenir une `hbase-site` classification qui spécifie l'emplacement dans Amazon S3 où les HBase données sont stockées à l'aide de la `hbase.rootdir` propriété. Il doit contenir une classification `hbase` qui spécifie `s3` à l'aide de la propriété `hbase.emr.storageMode`. L'exemple suivant illustre un JSON extrait de code contenant ces paramètres de configuration.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
    }
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3"
    }
  }
]
```

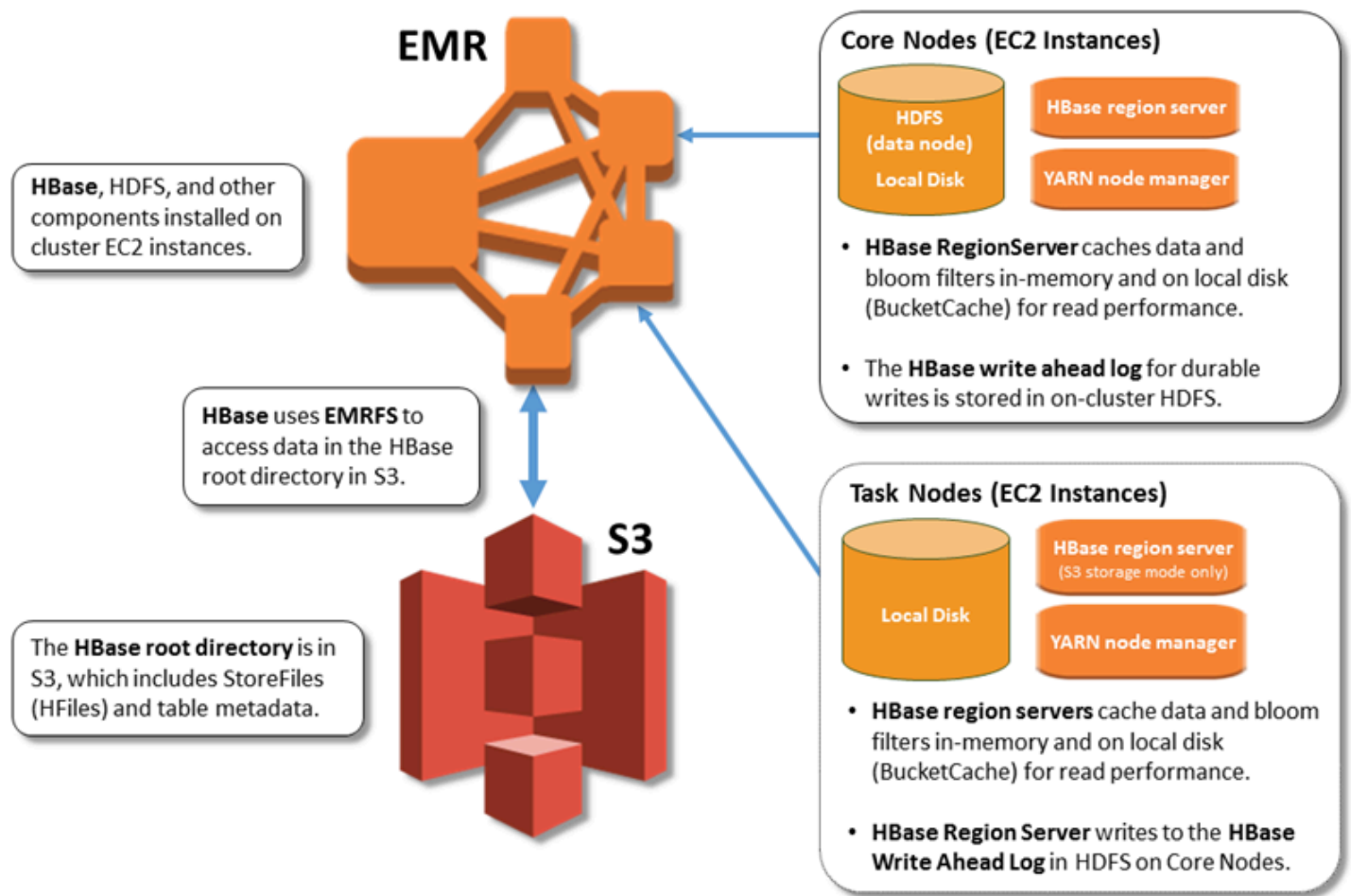
Pour plus d'informations HBase sur Amazon S3, consultez [HBasesur Amazon S3 \(mode de stockage Amazon S3\)](#). Pour plus d'informations sur les classifications, consultez [Configuration des applications](#).

HBasesur Amazon S3 (mode de stockage Amazon S3)

Lorsque vous exécutez HBase Amazon EMR version 5.2.0 ou ultérieure, vous pouvez l'activer HBase sur Amazon S3, ce qui offre les avantages suivants :

- Le répertoire HBase racine est stocké dans Amazon S3, y compris les fichiers de HBase stockage et les métadonnées des tables. Ces données sont persistantes en dehors du cluster, disponibles dans toutes les zones de EC2 disponibilité Amazon, et vous n'avez pas besoin de les récupérer à l'aide de snapshots ou d'autres méthodes.
- Avec les fichiers de stockage dans Amazon S3, vous pouvez dimensionner votre EMR cluster Amazon en fonction de vos besoins de calcul plutôt que de vos exigences en matière de données, avec une réplication multipliée par 3. HDFS
- À l'aide de EMR la version 5.7.0 ou ultérieure d'Amazon, vous pouvez configurer un cluster en lecture seule, qui vous permet de conserver des copies en lecture seule des données dans Amazon S3. Vous pouvez accéder aux données à partir du cluster de réplica en lecture pour réaliser des opérations de lecture simultanément et, dans le cas où le cluster principal devient indisponible.
- Dans Amazon EMR version 6.2.0 et versions ultérieures, le HFile suivi persistant utilise une table HBase système appelée `hbase:storefile` pour suivre directement les HFile chemins utilisés pour les opérations de lecture. Cette fonctionnalité est activée par défaut et ne nécessite pas de migration manuelle.

L'illustration suivante montre les HBase composants pertinents pour HBase Amazon S3.



Activation HBase sur Amazon S3

Vous pouvez l'activer HBase sur Amazon S3 à l'aide de la EMR console Amazon AWS CLI, du ou d'Amazon EMRAPI. La configuration est une option lors de la création du cluster. Lorsque vous utilisez la console, vous choisissez le paramètre à l'aide des Advanced options (Options avancées). Lorsque vous utilisez le AWS CLI, utilisez l'`--configurations` option pour fournir un objet JSON de configuration. Les propriétés de l'objet de configuration précisent le mode de stockage et l'emplacement du répertoire racine dans Amazon S3. L'emplacement Amazon S3 que vous spécifiez doit se trouver dans la même région que votre EMR cluster Amazon. Un seul cluster actif à la fois peut utiliser le même répertoire HBase racine dans Amazon S3. Pour les étapes de console et un exemple détaillé de création de cluster à l'aide du AWS CLI, voir. [Création d'un cluster avec HBase](#) Un exemple d'objet de configuration est présenté dans l'JSON extrait suivant.

```
{
  "Classification": "hbase-site",
```

```
"Properties": {
  "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
},
{
  "Classification": "hbase",
  "Properties": {
    "hbase.emr.storageMode": "s3"
  }
}
```

Note

Si vous utilisez un compartiment Amazon S3 comme formulaire `rootdirHBase`, vous devez ajouter une barre oblique à la fin du compartiment Amazon S3URI. Par exemple, vous devez utiliser `"hbase.rootdir: s3://my-bucket/"`, au lieu de `"hbase.rootdir: s3://my-bucket"`, pour éviter les problèmes.

Utilisation d'un cluster réplica en lecture

Après avoir configuré un cluster principal à l'aide d'HBaseAmazon S3, vous pouvez créer et configurer un cluster en lecture seule qui fournit un accès en lecture seule aux mêmes données que le cluster principal. Cela est utile lorsque vous avez besoin d'un accès simultané à des données de requête ou d'un accès ininterrompu si le cluster principal devient indisponible. La fonctionnalité de lecture et de réplication est disponible avec Amazon EMR version 5.7.0 et versions ultérieures.

Le cluster principal et le cluster de réplica en lecture sont définis de la même façon avec une différence significative. Les deux pointent le même emplacement `hbase.rootdir`. Cependant, la classification `hbase` pour le cluster de réplica en lecture comprend la propriété `"hbase.emr.readreplica.enabled": "true"`.

Par exemple, compte tenu de la JSON classification du cluster principal présentée plus haut dans cette rubrique, la configuration d'un cluster en lecture et réplique est la suivante :

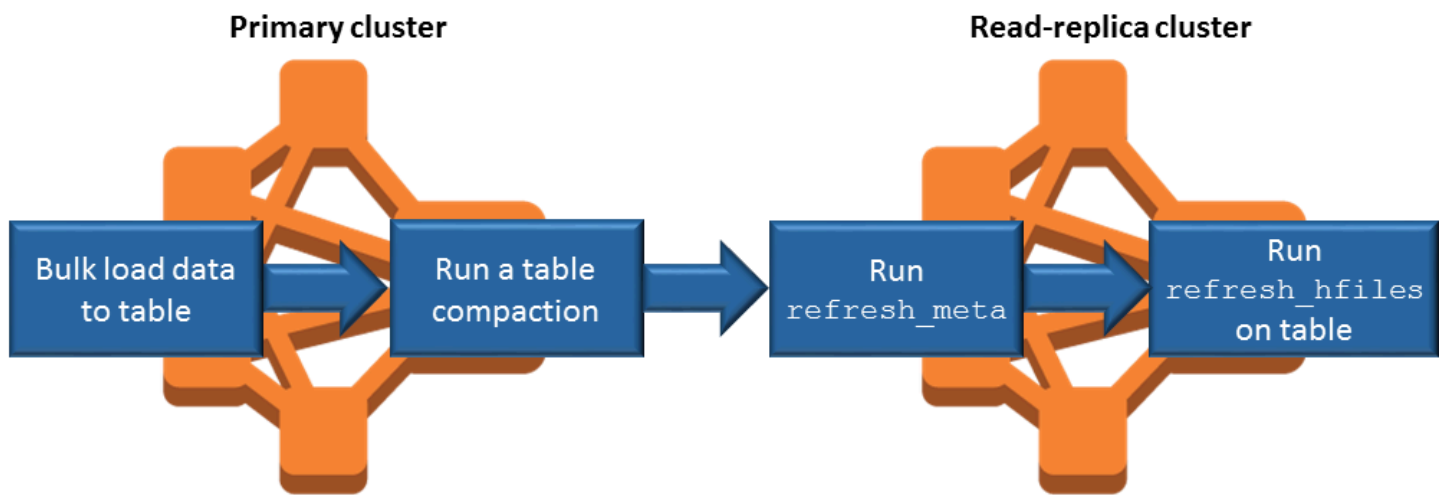
```
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"}
}
```

```
},
{
  "Classification": "hbase",
  "Properties": {
    "hbase.emr.storageMode": "s3",
    "hbase.emr.readreplica.enabled": "true"
  }
}
```

Synchronisation de réplica en lecture lors d'ajout de données

Étant donné que le réplica en lecture utilise HBase StoreFiles les métadonnées que le cluster principal écrit dans Amazon S3, le réplica en lecture est uniquement aussi actuel que le magasin de données Amazon S3. Les conseils suivants vous aident à minimiser le retard entre le cluster principal et la réplica en lecture lorsque vous écrivez des données.

- Chargez des données en bloc sur le cluster principal chaque fois que possible. Pour plus d'informations, consultez la section [Chargement groupé](#) dans la HBase documentation d'Apache.
- Un vidage qui écrit des fichiers de stockage dans Amazon S3 devrait être réalisé dès que possible après l'ajout des données. Procédez au vidage manuel ou configurez les paramètres de vidage pour minimiser le temps de retard.
- Si les compactages doivent s'exécuter automatiquement, exécutez un compactage manuel afin d'éviter les incohérences lors du déclenchement des compactages.
- Sur le cluster de lecture et de réplication, lorsque des métadonnées ont changé, par exemple en cas de division de HBase région ou de compactage, ou lorsque des tables sont ajoutées ou supprimées, exécutez la commande `refresh_meta`
- Sur le cluster de réplica en lecture, exécutez la commande `refresh_hfiles` lorsque des enregistrements sont ajoutés vers ou modifiés dans une table.



HFileSuivi permanent

HFileLe suivi permanent utilise une table HBase système appelée `hbase:storefile` pour suivre directement les HFile chemins utilisés pour les opérations de lecture. De nouveaux HFile chemins sont ajoutés à la table au fur et à mesure que des données supplémentaires sont ajoutées HBase. Cela supprime les opérations de renommage en tant que mécanisme de validation dans les HBase opérations de chemin d'écriture critiques et améliore le temps de restauration lors de l'ouverture d'une HBase région en lisant dans la table `hbase:storefile` système plutôt que dans la liste des répertoires du système de fichiers. Cette fonctionnalité est activée par défaut sur Amazon EMR version 6.2.0 et versions ultérieures, et ne nécessite aucune étape de migration manuelle.

Note

Le HFile suivi permanent à l'aide de la table système HBase Storefile ne prend pas en charge la fonctionnalité de réplification HBase régionale. Pour plus d'informations sur HBase la réplification régionale, voir Lectures à [haute disponibilité cohérentes avec la chronologie](#).

Désactiver le suivi permanent HFile

HFileLe suivi permanent est activé par défaut à partir de la EMR version 6.2.0 d'Amazon. Pour désactiver le HFile suivi permanent, spécifiez la modification de configuration suivante lors du lancement d'un cluster :

```
{
  "Classification": "hbase-site",
```

```
"Properties": {
  "hbase.storefile.tracking.persist.enabled":"false",

"hbase.hstore.engine.class":"org.apache.hadoop.hbase.regionserver.DefaultStoreEngine"
}
}
```

Note

Lors de la reconfiguration du EMR cluster Amazon, tous les groupes d'instances doivent être mis à jour.

Synchronisation manuelle de la table Storefile

La table Storefile est mise à jour au fur et à mesure que de nouvelles tables HFiles sont créées. Toutefois, si la table Storefile n'est plus synchronisée avec les fichiers de données pour une raison quelconque, les commandes suivantes peuvent être utilisées pour synchroniser manuellement les données :

Synchroniser la table Storefile dans une région en ligne :

```
hbase org.apache.hadoop.hbase.client.example.RefreshHFilesClient <table>
```

Synchroniser la table Storefile dans une région hors connexion :

- Supprimez la table Storefile znode.

```
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[<tableName>, hbase:namespace]
# The TableName exists in the list
echo "delete /hbase/storefile/loaded/<tableName>" | sudo -u hbase hbase zkcli
# Delete the Table ZNode
echo "ls /hbase/storefile/loaded" | sudo -u hbase hbase zkcli
[hbase:namespace]
```

- Attribuez la région (exécutez dans « hbase shell »).

```
hbase cli> assign '<region name>'
```

- Si l'assignation échoue.

```
hbase cli> disable '<table name>'
hbase cli> enable '<table name>'
```

Mise à l'échelle de la table Storefile

La table Storefile est divisée en quatre régions par défaut. Si la table Storefile est toujours soumise à une charge d'écriture importante, elle peut être divisée manuellement davantage.

Pour diviser une région chaude spécifique, utilisez la commande suivante (exécutée dans « hbase shell »).

```
hbase cli> split '<region name>'
```

Pour diviser la table, utilisez la commande suivante (exécutée dans « hbase shell »).

```
hbase cli> split 'hbase:storefile'
```

Considérations opérationnelles

HBase les serveurs régionaux sont utilisés BlockCache pour stocker les lectures de données en mémoire et BucketCache pour stocker les lectures de données sur le disque local. En outre, les serveurs régionaux stockent MemStore les écritures de données en mémoire et utilisent des journaux d'écriture anticipée pour stocker les écritures de données HDFS avant qu'elles ne soient écrites dans HBase StoreFiles Amazon S3. Les performances de lecture de votre cluster dépendent de la fréquence à laquelle un enregistrement peut être récupéré à partir des caches en mémoire ou sur le disque. Une erreur de cache entraîne la lecture de l'enregistrement StoreFile dans Amazon S3, ce qui présente une latence et un écart-type nettement supérieurs à ceux de la lecture depuis HDFS. De plus, les taux de demandes maximal pour Amazon S3 sont inférieurs à ce qui peut être réalisé à partir du cache local, la mise en cache de données peut donc être importante pour les charges de travail avec une forte densité de lectures. Pour plus d'informations sur les performances d'Amazon S3, consultez la section [Optimisation des performances](#) dans le Guide de l'utilisateur Amazon Simple Storage Service.

Pour améliorer les performances, nous vous recommandons de mettre en cache la plus grande partie possible de votre ensemble de données dans le stockage d'EC2instance. Comme il BucketCache utilise le stockage d'EC2instance du serveur régional, vous pouvez choisir un type d'EC2instance

avec un stockage d'instance suffisant et ajouter du EBS stockage Amazon pour répondre à la taille de cache requise. Vous pouvez également augmenter la BucketCache taille des magasins et EBS volumes d'instances attachés à l'aide de cette `hbase.bucketcache.size` propriété. Le paramètre par défaut est 8 192 MB.

En ce qui concerne les écritures, la fréquence des MemStore purges d'eau et le nombre de personnes StoreFiles présentes lors de compactages mineurs ou majeurs peuvent contribuer de manière significative à l'augmentation des temps de réponse des serveurs régionaux. Pour des performances optimales, pensez à augmenter la taille du MemStore purgeur et du multiplicateur de HRegion blocs, ce qui augmente le temps écoulé entre les compactages majeurs, mais augmente également le retard de cohérence si vous utilisez une réplique en lecture. Dans certains cas, vous pouvez obtenir de meilleures performances en utilisant des blocs de fichiers de plus grande taille (mais inférieurs à 5 Go) pour activer la fonctionnalité de téléchargement partitionné d'Amazon S3 dans EMRFS. La taille EMR de bloc par défaut d'Amazon est de 128 Mo. Pour de plus amples informations, veuillez consulter [HDFSconfiguration](#). Nous voyons rarement des clients dépasser une taille de bloc de 1 Go lorsque nous comparons les performances avec des vidages et des compactages. De plus, HBase les compactages et les serveurs régionaux fonctionnent de manière optimale lorsqu'il est StoreFiles nécessaire de compacter moins de serveurs.

Les tables peuvent mettre du temps pour arriver dans Amazon S3, car des répertoires volumineux doivent être renommés. Envisagez la désactivation des tables plutôt que leur suppression.

Il existe un HBase processus de nettoyage qui nettoie les anciens WAL fichiers et stocke les fichiers. Avec les EMR versions 5.17.0 et ultérieures d'Amazon, le nettoyeur est activé globalement et les propriétés de configuration suivantes peuvent être utilisées pour contrôler le comportement du nettoyeur.

Propriété de configuration	Valeur par défaut	Description
<code>hbase.regionserver.hfilecleaner.large.thread.count</code>	1	Le nombre de threads alloués au nettoyage a expiré en grande quantité HFiles.
<code>hbase.regionserver.hfilecleaner.small.thread.count</code>	1	Le nombre de threads alloués au nettoyage a expiré petit HFiles.

Propriété de configuration	Valeur par défaut	Description
<code>hbase.cleaner.scan.dir.concurrent.size</code>	Défini sur un quart de tous les noyaux disponibles.	Le nombre de threads à analyser dans les oldWALs répertoires.
<code>hbase.oldwals.cleaner.thread.size</code>	2	Le nombre de threads à nettoyer WALs sous le oldWALs répertoire.

Avec Amazon EMR 5.17.0 et versions antérieures, le fonctionnement du nettoyeur peut affecter les performances des requêtes lors de l'exécution de charges de travail importantes. Nous vous recommandons donc d'activer le nettoyeur uniquement pendant les heures creuses. Le nettoyeur possède les commandes HBase shell suivantes :

- `cleaner_chore_enabled` vérifie si l'appareil est activé.
- `cleaner_chore_run` exécute manuellement le nettoyeur pour supprimer les fichiers.
- `cleaner_chore_switch` active ou désactive le nettoyeur et le remet à son état précédent. Par exemple, `cleaner_chore_switch true` active le nettoyeur.

Propriétés pour HBase le réglage des performances sur Amazon S3

Les paramètres suivants peuvent être ajustés pour optimiser les performances de votre charge de travail lorsque vous l'utilisez HBase sur Amazon S3.

Propriété de configuration	Valeur par défaut	Description
<code>hbase.bucketcache.size</code>	8 192	La quantité d'espace disque, en Mo, réservée sur les stockages d'EC2instances Amazon sur le serveur régional et les EBS volumes de BucketCache stockage. Le paramètre s'applique à toutes les instances de serveur de

Propriété de configuration	Valeur par défaut	Description
		la région. Des BucketCache tailles plus grandes correspondent généralement à de meilleures performances
<code>hbase.hregion.memstore.flush.size</code>	134217728	La limite des données, en octets, à laquelle le vidage d'un memstore vers Amazon S3 est déclenché.
<code>hbase.hregion.memstore.block.multiplier</code>	4	Un multiplicateur qui détermine la limite MemStore supérieure à laquelle les mises à jour sont bloquées. Si les MemStore dépassent <code>hbase.hregion.memstore.flush.size</code> sont multipliés par cette valeur, les mises à jour sont bloquées. MemStore des bouffées d'eau et du compactage peuvent survenir pour débloquer les mises à jour.
<code>hbase.hstore.blockingStoreFiles</code>	10	Le nombre maximum de celles-ci peut exister dans un magasin avant StoreFiles que les mises à jour ne soient bloquées.
<code>hbase.hregion.max.filesize</code>	10737418240	La taille maximale d'une région avant que la région ne soit fractionnée.

Arrêt et restauration d'un cluster sans perte de données

Pour arrêter un EMR cluster Amazon sans perdre de données qui n'ont pas été écrites sur Amazon S3, vous devez vider votre MemStore cache d'Amazon S3 pour y écrire de nouveaux fichiers de boutique. Tout d'abord, vous devez désactiver toutes les tables. La configuration de l'étape suivante peut être utilisée lorsque vous ajoutez une étape au cluster. Pour plus d'informations, consultez la section [Travailler avec les étapes à l'aide de la console AWS CLI et](#) dans le guide EMR de gestion Amazon.

```
Name="Disable all tables",Jar="command-runner.jar",Args=["/bin/bash","/usr/lib/hbase/bin/disable_all_tables.sh"]
```

Vous pouvez également exécuter la commande bash suivante directement.

```
bash /usr/lib/hbase/bin/disable_all_tables.sh
```

Après avoir désactivé toutes les tables, videz-les à l'hbase :met aaide du HBase shell et de la commande suivante.

```
flush 'hbase:meta'
```

Vous pouvez ensuite exécuter un script shell fourni sur le EMR cluster Amazon pour vider le MemStore cache. Vous pouvez l'ajouter soit comme une étape, soit en l'exécutant directement à l'aide d' AWS CLI sur le cluster. Le script désactive toutes les HBase tables, ce qui MemStore entraîne le transfert du serveur de chaque région vers Amazon S3. Si le script se termine avec succès, les données persistent dans Amazon S3 et le cluster peut être arrêté.

Pour redémarrer un cluster avec les mêmes HBase données, spécifiez le même emplacement Amazon S3 que le cluster précédent dans AWS Management Console ou à l'aide de la propriété `hbase.rootdir` de configuration.

Journaux d'écriture anticipée (WAL) pour Amazon EMR

Avec Amazon EMR 6.15 et versions ultérieures, vous pouvez écrire vos journaux d'HBaseécriture anticipée Apache (WAL) sur Amazon. EMR WAL Dans les EMR versions inférieures d'Amazon, lorsque vous créez un cluster avec l'option HBasesur Amazon S3, WAL c'est le seul HBase composant Apache qui est stocké sur le disque local pour les clusters, et vous pouvez stocker

d'autres composants tels que le répertoire racine, les fichiers de stockage (HFiles), les métadonnées des tables et les données sur Amazon S3.

Vous pouvez utiliser Amazon EMR WAL pour récupérer des données qui n'ont pas été transférées vers Amazon S3. Pour sauvegarder l'intégralité de vos HBase clusters, choisissez d'utiliser le EMR WAL service Amazon. Dans les coulisses, `RegionServer` écrit vos journaux d'HBaseécriture anticipée (WAL) dans le pour WAL Amazon. EMR

Si votre cluster ou l'AZ ne fonctionne pas correctement ou n'est plus disponible, vous pouvez créer un nouveau cluster, le faire pointer vers le même répertoire racine S3 et le même EMR WAL espace de travail Amazon, et récupérer automatiquement les données en quelques WAL minutes. Pour de plus amples informations, veuillez consulter [Restauration depuis Amazon EMR WAL](#).

Note

Amazon EMR conserve votre journal d'écriture anticipée et ses données pendant 30 jours à compter de la création de votre cluster. Au bout de 30 jours, Amazon supprime EMR automatiquement votre Amazon EMR WAL et ses données. Toutefois, si vous lancez un nouveau cluster WAL activé à partir du même répertoire racine S3, vous pouvez prolonger l'utilisation de votre cluster WAL pendant 30 jours à compter de la date de lancement du nouveau cluster. Amazon EMR continuera à nettoyer toutes WAL les données du premier cluster après la période initiale de 30 jours. Pour de plus amples informations, veuillez consulter [Restauration depuis Amazon EMR WAL](#).

Les sections suivantes décrivent comment configurer et utiliser Amazon EMR WAL avec votre EMR cluster HBase activé.

Rubriques

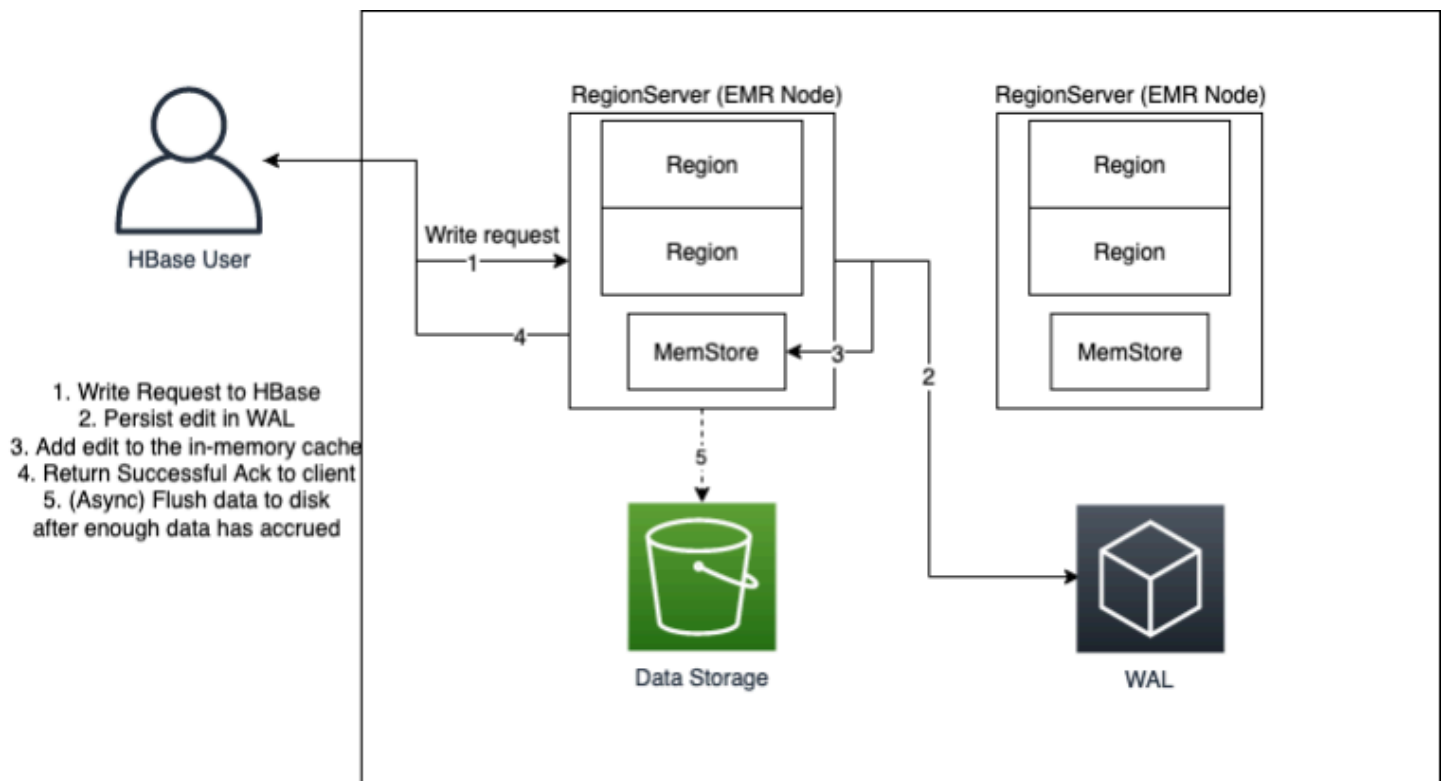
- [Espaces de EMR WAL travail Amazon](#)
- [Autorisations requises pour Amazon EMR WAL](#)
- [Activation d'Amazon EMR WAL](#)
- [Restauration depuis Amazon EMR WAL](#)
- [Utilisation de configurations de sécurité avec Amazon EMR WAL](#)
- [Accédez à Amazon EMR WAL via AWS PrivateLink](#)
- [Comprendre les EMR WAL tarifs et les statistiques d'Amazon](#)

- [Marquage WAL des espaces de travail](#)
- [Considérations et régions pour Amazon EMR WAL](#)
- [CLI Référence Amazon EMR WAL \(EMRWAL\)](#)

Espaces de EMR WAL travail Amazon

Amazon EMR WAL ajoute le concept d'WAL espaces de travail. Un WAL espace de travail est un conteneur logique de WALs. Chaque journal écrit à l'avance sur Amazon EMR WAL est encapsulé dans un espace de travail. WAL Un EMR cluster écrit WALs dans un seul WAL espace de travail que vous configurez au lancement du cluster, ou dans le `defaultWALworkspace` si vous ne spécifiez aucun espace de travail. WAL les espaces de travail ne sont liés à aucune HBase terminologie existante telle que les espaces de noms.

Vous pouvez utiliser WAL les espaces de travail pour limiter les EMR WAL IAM autorisations Amazon afin d'inclure uniquement les espaces de travail auxquels le cluster doit accéder. Vous pouvez également étiqueter votre WAL espace de travail pour un contrôle d'accès basé sur des balises. Pour plus d'informations sur le balisage, consultez [Marquage WAL des espaces de travail](#).



Autorisations requises pour Amazon EMR WAL

Pour que votre cluster puisse se connecter à Amazon EMR WAL, le profil d'instance du cluster nécessite certaines IAM autorisations :

- Amazon EMR WAL utilise le rôle [AWSServiceRoleForEMRWAL](#) lié à un service pour récupérer le statut d'un cluster. Amazon crée EMR automatiquement ce rôle lié au service lorsque vous créez un WAL espace de travail, ou HBase créera le rôle lié au service lorsque vous configurez un espace de travail pour Amazon EMR WAL et que le rôle lié au service n'existe pas encore.

Avant de pouvoir activer Amazon EMR WAL pour un cluster, vous devez configurer les autorisations afin de permettre la création automatique du rôle [AWSServiceRoleForEMRWAL](#) lié au service. Pour plus d'informations et un exemple d'instruction qui ajoute cette fonctionnalité, consultez la section [Utilisation de rôles liés à un service pour la journalisation en écriture anticipée](#).

- Amazon EMR WAL utilisant HBase Write Ahead Log (WAL), vos clusters doivent utiliser HBaseWAL. Les IAM autorisations minimales dont vous avez besoin pour exécuter sont les suivantes HBase. Ajoutez les éléments suivants à la politique d'autorisation de votre profil d'instance :

```
emrwal:DeleteWal
emrwal:CreateWal
emrwal:CreateWorkspace
emrwal:AppendEdit
emrwal:ReplayEdits
emrwal:GetCurrentWalTime
emrwal:CompleteWalFlush
```

Note

Si vous limitez les autorisations EMR WAL d'Amazon à l'ensemble minimal, certaines [EMRWALCLI](#) commandes ne disposeront pas des autorisations nécessaires pour s'exécuter.

Activation d'Amazon EMR WAL

Suivez les étapes suivantes pour activer l'écriture sur Amazon EMR WAL lorsque vous créez un cluster avec le AWS Command Line Interface.

Note

Vous ne pouvez pas activer Amazon EMR WAL pour un cluster déjà en cours d'exécution, et vous ne pouvez pas lancer deux clusters avec le même répertoire racine S3. Pour de plus amples informations, veuillez consulter [Considérations et régions pour Amazon EMR WAL](#).

1. Avant de créer un cluster EMR WAL compatible avec Amazon, vous devez ajouter les autorisations requises au profil d'instance que vous prévoyez d'utiliser avec votre cluster. Pour de plus amples informations, veuillez consulter [Autorisations requises pour Amazon EMR WAL](#).
2. Créez un cluster à partir du AWS CLI. Utilisez `--configurations` cette option pour fournir un objet JSON de configuration qui spécifie la `hbase.emr.wal.enabled` propriété, comme illustré dans l'exemple ci-dessous.
 - Spécifiez le mode de stockage et l'emplacement du répertoire racine dans Amazon S3. L'emplacement Amazon S3 que vous spécifiez doit se trouver dans la même région que votre EMR cluster, mais un seul cluster actif peut utiliser le même répertoire HBase racine dans S3 à la fois.
 - Créez votre cluster avec la configuration des groupes d'instances. Vous ne pouvez pas utiliser Amazon EMR WAL avec la configuration des flottes d'instances. Pour plus d'informations sur la création de clusters avec des groupes d'instances, consultez [Configurer des groupes d'instances uniformes](#) dans le Amazon EMR Management Guide.
 - Pour connaître les étapes de création d'un cluster sur console, ainsi qu'un `create-cluster` exemple détaillé utilisant le AWS CLI, voir [Création d'un cluster avec HBase](#).
3. WALPour activer le nouveau cluster, définissez la `hbase.emr.wal.enabled` propriété sur `true`. La commande suivante contient un JSON extrait de code avec un exemple d'objet de configuration.

```
aws emr create-cluster --name "hbasewal" --release-label emr-6.x.y \  
--applications Name=HBase --use-default-roles --ec2-attributes KeyName=myKey \  
--instance-type m6i.xlarge --instance-count 1 --configurations hbase.json  
$cat hbase.json  
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"  
    }  
  }  
]
```

```

    },
    {
      "Classification": "hbase",
      "Properties": {
        "hbase.emr.storageMode": "s3",
        "hbase.emr.wal.enabled": "true"
      }
    }
  ]

```

Lorsqu'il HBase est en ligne sur le cluster nouvellement créé, HBase il écrit automatiquement WAL des données sur Amazon EMR WAL et utilise l'Amazon EMR WAL à des fins de récupération.

Exemple 1 : Création d'un EMR cluster utilisant Amazon EMR WAL

```

[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
    }
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.wal.enabled": "true"
    }
  }
]

```

Exemple 2 : Création d'un EMR cluster avec un WAL espace de travail personnalisé

```

[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore",
      "emr.wal.workspace": "customWorkspaceName"
    }
  },
  {

```



```
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.wal.enabled": "true"
    }
  }
]
```

Restauration depuis Amazon EMR WAL

Étant donné que l'Amazon EMR WAL de votre cluster d'origine est conservé pendant 30 jours, vous pouvez le restaurer et le réutiliser WAL pour un cluster nouvellement créé au cours de cette période de 30 jours. Lorsque vous lancez un nouveau cluster à partir du même répertoire racine S3, le délai de 30 jours recommence à compter de l'heure de lancement du nouveau cluster, à condition que la période de 30 jours précédente n'ait pas expiré.

Utilisez la procédure suivante pour restaurer un cluster existant WAL avec un nouveau cluster. Ce processus suppose que vous avez créé votre cluster d'origine avec Amazon EMR WAL activé.

1. Dans les 30 jours suivant la création d'un cluster WAL activé, créez un nouveau cluster Région AWS identique au cluster d'origine. Le nouveau cluster peut se trouver dans la même AZ ou dans une autre AZ au sein de la même région que celle dans laquelle le cluster d'origine a été créé.

Configurez les propriétés de l'objet pour spécifier le mode de stockage et l'emplacement du répertoire racine dans Amazon S3. L'emplacement Amazon S3 que vous spécifiez doit se trouver dans la même région que votre EMR cluster, mais un seul cluster actif peut utiliser le même répertoire HBase racine dans S3 à la fois.

Pour connaître les étapes de création d'un cluster sur console, ainsi qu'un `create-cluster` exemple détaillé utilisant le AWS CLI, voir [Création d'un cluster avec HBase](#).

2. Pour utiliser l'Amazon existant EMR WAL pour le nouveau cluster, définissez la `hbase.emr.wal.enabled` propriété sur `true`. L'JSON extrait suivant montre un exemple d'objet de configuration.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
```

```
        "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
    }
},
{
    "Classification": "hbase",
    "Properties": {
        "hbase.emr.storageMode": "s3",
        "hbase.emr.wal.enabled": "true"
    }
}
]
```

Utilisation de configurations de sécurité avec Amazon EMR WAL

Amazon chiffre EMR automatiquement les données en transit entre votre cluster et le EMR WAL service Amazon, ainsi que les données inactives sur Amazon EMRWAL. Pour plus d'informations, consultez [Encryption at rest for Amazon EMR WAL](#). Vous pouvez également utiliser une configuration de sécurité pour importer vos propres clés depuis le service AWS Key Management Service (KMS) et chiffrer les données que vous stockez sur Amazon EMRWAL.

Utilisez l'une des méthodes suivantes pour sélectionner une configuration de sécurité lorsque vous créez un cluster :

Console

À partir de AWS Management Console, spécifiez la configuration sous Configuration de sécurité et paire de EC2 clés.

Security configuration and EC2 key pair - *optional* [Info](#)

Security configuration

Select your cluster encryption, authentication, authorization, and instance metadata service settings.



CLI

À partir de AWS CLI, définissez le `--security-configuration` paramètre lorsque vous utilisez la commande [create-cluster](#).

Pour plus d'informations, consultez [Encryption at rest for Amazon EMR WAL](#) et [Utiliser des configurations de sécurité pour configurer la sécurité du cluster](#) dans le guide EMR de gestion Amazon.

Pour plus d'informations relatives à la sécurité WAL, voir [Utilisation de rôles liés à un service](#) pour la journalisation anticipée.

Accédez à Amazon EMR WAL via AWS PrivateLink

Si vous souhaitez conserver votre connexion au sein du AWS réseau, Amazon EMR WAL propose une AWS PrivateLink assistance. Pour le configurer AWS PrivateLink, utilisez le AWS Management Console ou AWS Command Line Interface (AWS CLI) pour créer un point de VPC terminaison d'interface qui se connecte à Amazon EMRWAL. Pour plus d'informations, consultez la section [Accès à un AWS service à l'aide d'un point de VPC terminaison d'interface](#) dans le AWS PrivateLink Guide.

La procédure de base est la suivante :

1. Utilisez la VPC console Amazon pour [créer un VPC point de terminaison](#). Sélectionnez Endpoints, puis Create endpoint.
2. Conservez la catégorie Service en tant que AWS services.
3. Dans la barre de recherche du panneau Services, tapez **emrwal**, puis sélectionnez le service intitulé `com.amazonaws.region.emrwal.prod`.
4. Sélectionnez votre VPC et enregistrez le point de terminaison. Assurez-vous d'associer les mêmes groupes de sécurité au point de VPC terminaison que celui que vous attachez au EMR cluster.
5. Si vous le souhaitez, vous pouvez désormais activer les DNS noms d'hôte privés pour votre nouveau point de terminaison. Définissez Enable DNS hostnames et Enable DNS Support sur `true` pour votre VPC. Sélectionnez ensuite votre identifiant de point de terminaison, choisissez Modifier VPC les paramètres dans le menu Actions et activez les DNS noms privés.
 - Les DNS noms d'hôtes privés du point de terminaison suivront le format `prod.emrwal.region.amazonaws.com`.
 - Si vous n'activez pas les DNS noms d'hôte privés, Amazon vous VPC fournit un nom de point de DNS terminaison au format `endpointID.prod.emrwal.region.vpce.amazonaws.com` suivant.

6. Pour utiliser votre AWS PrivateLink point de terminaison, modifiez la `emr.wal.client.endpoint` configuration lorsque vous créez votre [cluster EMR WAL compatible Amazon](#), comme indiqué dans l'exemple suivant :

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore",
      "emr.wal.workspace": "customWorkspaceName",
      "emr.wal.client.endpoint": "https://prod.emrwal.region.amazonaws.com"
    }
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.wal.enabled": "true"
    }
  }
]
```

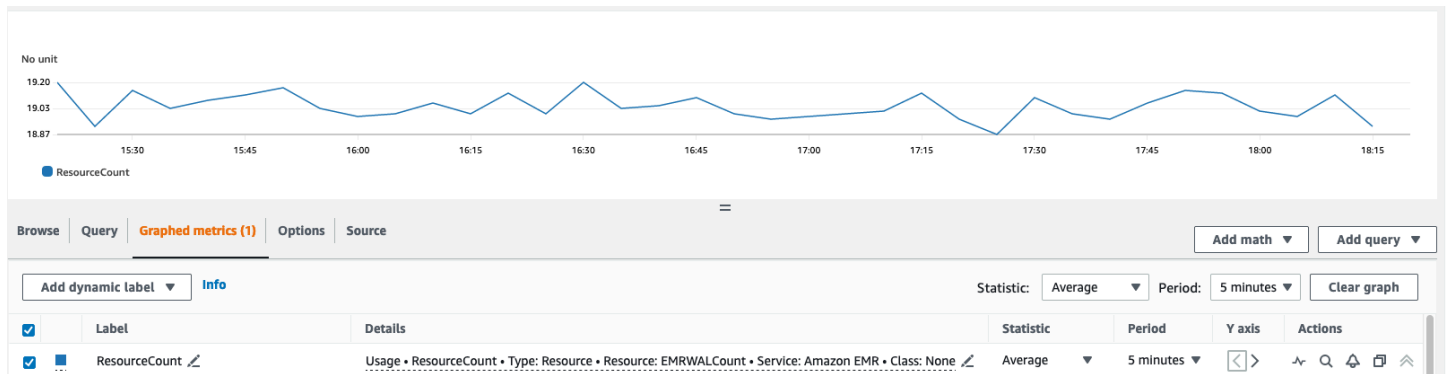
Vous pouvez également utiliser VPCE la politique pour autoriser ou restreindre l'accès à Amazon EMR WALAPIs. Pour plus d'informations, consultez la section [Contrôler l'accès aux points de VPC terminaison à l'aide des politiques relatives aux points de terminaison](#) dans le AWS PrivateLink Guide.

Comprendre les EMR WAL tarifs et les statistiques d'Amazon

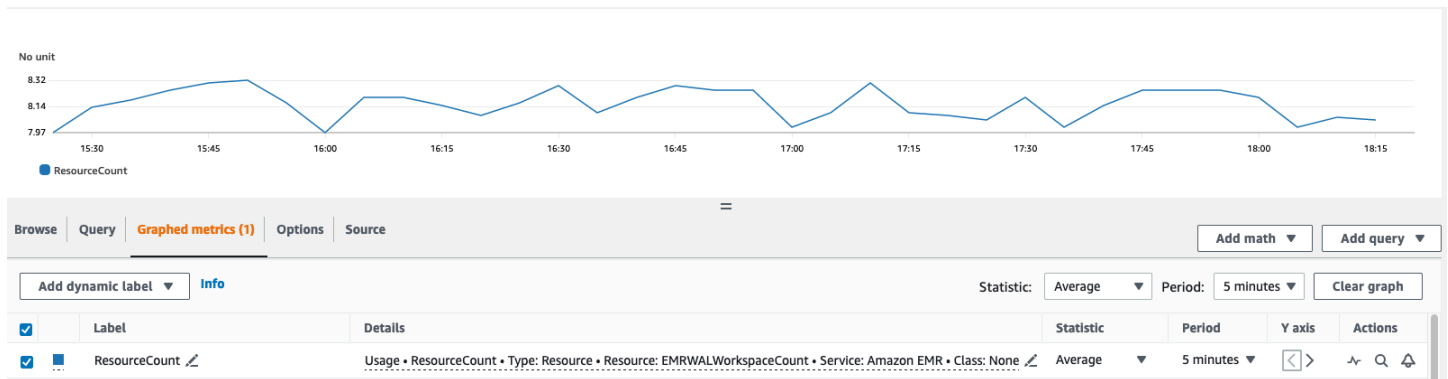
Unité de facturation des fonctionnalités de base	Détails
EMR-WAL-Read-GiB	API calls to read data from your table are billed as ReadRequestGiB. This includes Get and Scan operations. Reads are charged based on the sizes of the read items. Amazon EMR bills at a minimum of 1 byte. For example, if you read a 1234.12 bytes item, you're

Unité de facturation des fonctionnalités de base	Détails
	<p>charged for 1235 bytes. Reads are aggregated every hour for billing and shown as GiBs.</p>
EMR-WAL-Write-GiB	<p>API calls to write data from your table are billed as Write-GiB. This includes Put operations. Writes are charged based on the sizes of the written items. Amazon EMR bills at a minimum of 1 byte. For example, if you write a 1234.12 bytes item, you're charged for 1235 bytes. Writes are aggregated every hour for billing and shown as GiBs.</p>
EMR-WAL-WALHours	<p>Le numéro WALs que vous enregistrez sur le service est facturé comme <code>EMR-WAL-WALHours</code>. Amazon en EMR crée un WAL par HBase région. Par exemple, si vous créez 20 HBase tables, y compris des tables système, et que chaque table comporte deux HBase régions, vous utilisez 28 800 WAL heures, calculées comme suit :</p> <div style="border: 1px solid #ccc; border-radius: 10px; padding: 10px; margin-top: 10px;"> <pre> 20 tables x 2 Regions per table x 1 WAL per Region x 30 days x 24 hours ----- 28,800 EMR-WAL-WALHours </pre> </div>

Exemple **EMRWALCount** :



Exemple EMRWALWorkspaceCount :



Marquage WAL des espaces de travail

Vous pouvez ajouter des balises à un espace de travail lorsque vous créez un nouvel espace de travail et vous pouvez ajouter, supprimer ou répertorier des balises d'un espace de travail actif pour un cluster en cours d'exécution. Vous ne pouvez pas étiqueter les ressources individuelles de l'espace de travail, ni mettre à jour les balises existantes. Supprimez plutôt les balises indésirables de l'espace de travail et remplacez-les.

Vous pouvez baliser des espaces de travail à partir du EMRWAL CLI. Pour obtenir la liste des EMRWAL CLI commandes permettant de baliser les espaces de travail, consultez. [CLIRéférence Amazon EMR WAL \(EMRWAL\)](#)

L'exemple de IAM politique suivant illustre un scénario qui autorise les CRUDL opérations d'espace de travail uniquement avec la clé `resource_tag_allow_test_key` et la valeur `resource_tag_allow_test_value` de balisage appropriées :

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
```

```

    "Action": [
      "EMRWAL:*"
    ],
    "Effect": "Allow",
    "Resource": [
      "*",
      "*"
    ],
    "Condition": {
      "StringEquals": {
        "aws:ResourceTag/resource_tag_allow_test_key": [
          "resource_tag_allow_test_value"
        ]
      }
    }
  }
]
}

```

Pour vérifier que la balise est désormais requise pour les opérations de l'espace de travail, utilisez la [CLIRéférence Amazon EMR WAL \(EMRWAL\) listTagsForResource](#) commande `tagAllowResourceTag` pour activer l'espace de travail avec la balise de ressource souhaitée. Si vous avez correctement configuré la condition, la commande aboutira.

```

emrwal listTagsForResource -r us-east-1 -arn arn:aws:emrwal:us-east-1:arn:workspace/
tagAllowResourceTag
Tag(Key=resource_tag_allow_test_key, Value=resource_tag_allow_test_value)

```

Considérations et régions pour Amazon EMR WAL

Considérations pour Amazon EMR WAL

La liste suivante décrit les principales considérations et limites d'Amazon EMR WAL :

- Amazon EMR WAL peut être utilisé avec les EMR versions 6.15.0 et supérieures d'Amazon.
- Amazon EMR WAL est un service payant et optionnel. Vous payez pour ce que vous utilisez : lecture, écriture et stockage de données. Pour plus d'informations, consultez [Comprendre les EMR WAL tarifs et les statistiques d'Amazon](#) la page de [EMRtarification d'Amazon](#).
- Amazon EMR WAL utilise HBase Write Ahead Log (WAL). Pour utiliser Amazon EMRWAL, vos clusters doivent utiliser HBaseWAL.

- Pour activer Amazon EMR WAL lorsque vous créez un cluster, vous devez disposer des autorisations de rôle requises. Pour plus d'informations, consultez la section [Utilisation de rôles liés à un service pour la journalisation en écriture anticipée](#).
- Vous devez activer Amazon EMR WAL lorsque vous créez le cluster avec le AWS Management Console, ou AWS CLI API, et vous devez utiliser la configuration des groupes d'instances. Vous ne pouvez pas activer Amazon EMR WAL dans un cluster en cours d'exécution si vous n'avez pas créé le cluster avec Amazon EMRWAL. Vous ne pouvez pas non plus modifier les hbase-site configurations pour activer Amazon EMR WAL dans un cluster en cours d'exécution.
- Vous ne pouvez activer Amazon que EMR WAL sur des clusters qui utilisent Amazon S3 comme répertoire racine.
- Vous ne pouvez pas avoir plusieurs clusters actifs sur le même répertoire HBase racine dans Amazon S3.
- Vous ne pouvez pas activer Amazon EMR WAL sur des clusters Read Replica.
- WAL est répliqué dans les zones de disponibilité du service géré.
- WAL dure plus longtemps que le cluster et reste disponible pour le cluster suivant.
- Vous ne pouvez pas désactiver Amazon EMR WAL lors du lancement ou lorsque votre cluster est opérationnel (en cours d'exécution).
- Pour plus d'informations sur les limites de l'espace de travail WAL et les limites de l'espace de travail, consultez [Amazon EMR Endpoints and Quotas](#).

Disponibilité par région pour Amazon EMR WAL

Le EMR WAL service Amazon est disponible dans les pays suivants Régions AWS :

- `ap-northeast-1` – Asie-Pacifique (Tokyo)
- `ap-southeast-1` – Asie-Pacifique (Singapour)
- `ap-south-1` – Asie-Pacifique (Mumbai)
- `ap-southeast-2` – Asie-Pacifique (Sydney)
- `eu-central-1` – Europe (Francfort)
- `eu-north-1` – Europe (Stockholm)
- `eu-west-1` – Europe (Irlande)
- `us-east-1` – USA Est (Virginie du Nord)
- `us-east-2` – USA Est (Ohio)

- us-west-2 – USA Ouest (Oregon)
- ap-northeast-2 – Asie-Pacifique (Séoul)

CLI Référence Amazon EMR WAL (EMRWAL)

L'interface de ligne de commande EMRWAL (EMRWAL CLI) est un outil unifié permettant de gérer votre journal d'écriture anticipée (WAL) pour Amazon EMR. Il est livré avec EMR des clusters lorsque vous les activez WAL au moment de créer un cluster. Pour plus d'informations sur l'activation WAL, consultez [Journaux d'écriture anticipée \(WAL\) pour Amazon EMR](#).

EMRWAL CLI inclut les commandes suivantes :

Rubriques

- [createWorkspace](#)
- [deleteWal](#)
- [deleteWorkspace](#)
- [listTagsForResource](#)
- [listWals](#)
- [listWorkspaces](#)
- [tagResource](#)
- [untagResource](#)

createWorkspace

La createWorkspace commande crée un nouvel espace de EMR WAL travail Amazon.

Utilisation :

```
emrwal createWorkspace [-tags <tags>] [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

Exemple :

```
aws emr createWorkspace -w examplews
```

deleteWal

La deleteWals commande supprime l'Amazon EMR WAL que vous spécifiez.

Utilisation :

```
emrwal deleteWal [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>]
[-n <walName>] [-N <fullName>] [-R] [-m] [-h]
```

Exemple :

```
emrwal deleteWal -w examples -p hbasetable -n examplewal
```

deleteWorkspace

La deleteWorkspace commande supprime l'EMRWALespace de travail Amazon que vous spécifiez.

Utilisation :

```
emrwal deleteWorkspace [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

Exemple :

```
emrwal deleteWorkspace -w examples
```

listTagsForResource

La listTagsForResource commande répertorie toutes les paires clé-valeur que vous spécifiez pour l'EMRWALespace de travail Amazon.

Utilisation :

```
emrwal listTagsForResource -arn {resource-arn} [-e {endpoint}] [-r {Region}] [-h]
```

Exemple :

```
emrwal listTagsForResource -arn arn:aws:emrwal::1234567891234:workspace/examples
```

listWals

La `listWals` commande répertorie tous les Amazon présents EMR WALs dans l'espace de travail que vous spécifiez.

Utilisation :

```
emrwal listWals [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>] [-M {integer}] [-h]
```

Exemple :

```
emrwal listWals -w examples
```

listWorkspaces

La `listWorkspaces` commande répertorie tous les EMR WAL espaces de travail Amazon mis à votre disposition.

Utilisation :

```
emrwal listWorkspaces [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}] [-r {Region}] [-M {integer}] [-h]
```

Exemple :

```
emrwal listWorkspaces
```

tagResource

La `tagResource` commande attribue une ou plusieurs paires clé-valeur à l'EMRWALespace de travail Amazon que vous spécifiez.

Utilisation :

```
emrwal tagResource -arn {resource-arn} -tags <tags> [-e {endpoint}] [-r {Region}] [-h]
```

Exemple :

```
emrwal tagResource -arn arn:aws:emrwal::1234567891234:workspace/examples -  
tags tag_key=tag_value
```

untagResource

La untagResource commande annule l'attribution d'une ou de plusieurs paires clé-valeur à l'EMRWorkspace de travail Amazon que vous spécifiez.

Utilisation :

```
emrwal untagResource -arn {resource-arn} -tagKeys <tagKeys> [-e {endpoint}] [-  
r {Region}] [-h]
```

Exemple :

```
emrwal untagResource -arn arn:aws:emrwal::1234567891234:workspace/examples -  
tagKeys tag_key
```

Utilisation de la HBase coque

Après avoir créé un HBase cluster, l'étape suivante consiste à vous HBase y connecter afin de pouvoir commencer à lire et à écrire des données (les écritures de données ne sont pas prises en charge sur un cluster en lecture-réplique). Vous pouvez utiliser le [HBaseshell](#) pour tester les commandes.

Pour ouvrir la HBase coque

1. SSH à utiliser pour se connecter au serveur principal du HBase cluster. Pour plus d'informations sur la façon de se connecter au nœud EMR principal Amazon à l'aide de SSH, consultez [la section Se connecter au nœud principal SSH à l'aide](#) du guide EMR de gestion Amazon.
2. Exécutez `hbase shell`. Le HBase shell s'ouvre avec un message similaire à ce qui suit.

```
hbase(main):001:0>
```

Vous pouvez exécuter des commandes HBase shell à partir de l'invite. Pour plus d'informations sur les commandes shell et sur la façon de les appeler, tapez `help` à l'HBaseinvite et appuyez sur Entrée.

Créer une table

La commande suivante crée une table nommée « t1 », dotée d'une famille de colonnes unique nommée « f1 ».

```
hbase(main):001:0>create 't1', 'f1'
```

Placement d'une valeur

La commande suivante place la valeur « v1 » pour la ligne « r1 » dans la table « t1 » et la colonne « f1 ».

```
hbase(main):001:0>put 't1', 'r1', 'f1:col1', 'v1'
```

Obtention d'une valeur

La commande suivante obtient les valeurs pour la ligne « r1 » dans la table « t1 ».

```
hbase(main):001:0>get 't1', 'r1'
```

Supprimer une table

La commande suivante supprime et supprime le tableau « t1 ».

```
hbase(main):001:0>drop 'ns1:t1',false
```

La valeur booléenne indique si vous souhaitez archiver ou non votre table. Vous pouvez donc la définir sur `true` si vous souhaitez l'enregistrer. Vous pouvez également exécuter `drop 'ns1:t1'` sans booléen pour archiver la table.

Accédez aux HBase tables avec Hive

HBase et [Apache Hive](#) sont étroitement intégrés, ce qui vous permet d'exécuter des charges de travail de traitement massivement parallèles directement sur les données stockées dans HBase. Pour utiliser Hive avec HBase, vous pouvez généralement les lancer sur le même cluster. Vous pouvez toutefois lancer Hive et HBase sur des clusters distincts. Exécuter HBase et Hive séparément sur

différents clusters peut améliorer les performances, car cela permet à chaque application d'utiliser les ressources du cluster de manière plus efficace.

Les procédures suivantes montrent comment se connecter à un cluster HBase à l'aide de Hive.

Note

Vous ne pouvez connecter un cluster Hive qu'à un seul HBase cluster.

Pour connecter Hive à HBase

1. Créez des clusters séparés avec Hive et HBase installés ou créez un seul cluster avec les deux HBase et Hive installés.
2. Si vous utilisez des clusters distincts, modifiez vos groupes de sécurité afin que HBase les ports Hive soient ouverts entre ces deux nœuds principaux.
3. SSH à utiliser pour se connecter au nœud principal du cluster sur lequel Hive est installé. Pour plus d'informations, consultez [Se connecter au nœud principal SSH à l'aide](#) de l'Amazon EMR Management Guide.
4. Lancez le shell Hive à l'aide de la commande suivante.

```
hive
```

5. (Facultatif) Vous n'avez pas besoin de le faire si Hive HBase et Hive se trouvent sur le même cluster. Connectez le HBase client de votre cluster Hive au HBase cluster qui contient vos données. Dans l'exemple suivant, *public-DNS-name* est remplacé par le DNS nom public du nœud principal du HBase cluster, par exemple `ec2-50-19-76-67.compute-1.amazonaws.com`.

```
set hbase.zookeeper.quorum=public-DNS-name;
```

6. Exécutez les requêtes Hive sur vos HBase données comme vous le souhaitez ou consultez la procédure suivante.

Pour accéder aux HBase données depuis Hive

- Une fois la connexion établie entre le Hive et les HBase clusters (comme indiqué dans la procédure précédente), vous pouvez accéder aux données stockées sur le HBase cluster en créant une table externe dans Hive.

L'exemple suivant, lorsqu'il est exécuté à partir de l'invite Hive sur le nœud principal, crée une table externe qui référence les données stockées dans une HBase table appelée `inputTable`. Vous pouvez ensuite faire référence `inputTable` dans les instructions Hive pour interroger et modifier les données stockées dans le HBase cluster.

```
set hbase.zookeeper.quorum=ec2-107-21-163-157.compute-1.amazonaws.com;

create external table inputTable (key string, value string)
  stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
  with serdeproperties ("hbase.columns.mapping" = ":key,f1:col1")
  tblproperties ("hbase.table.name" = "t1");

select count(key) from inputTable ;
```

Pour un cas d'utilisation plus avancé et un exemple de combinaison HBase avec Hive, consultez le billet de blog sur le AWS Big Data, [Combine No SQL et massively parallel analytics using Apache HBase and Apache Hive on Amazon](#). EMR

Utilisation de HBase snapshots

HBase utilise une fonctionnalité de [capture instantanée](#) intégrée pour créer des sauvegardes légères de tables. Dans les EMR clusters, ces sauvegardes peuvent être exportées vers Amazon S3 à l'aide de EMRFS. Vous pouvez créer un instantané sur le nœud principal à l'aide du HBase shell. Cette rubrique explique comment exécuter ces commandes de manière interactive avec le shell ou en effectuant une étape à l'command-runner.jar aide du AWS CLI ou AWS SDK for Java. Pour plus d'informations sur les autres types de HBase sauvegardes, consultez la section [HBasesauvegarde](#) dans la HBase documentation.

Création d'un instantané à l'aide d'une table

```
hbase snapshot create -n snapshotName -t tableName
```

En utilisant le fichier `command-runner.jar` depuis AWS CLI :

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \  
--steps Name="HBase Shell Step",Jar="command-runner.jar",\  
Args=[ "hbase", "snapshot", "create", "-n", "snapshotName", "-t", "tableName"]
```

AWS SDK for Java

```
HadoopJarStepConfig hbaseSnapshotConf = new HadoopJarStepConfig()  
.withJar("command-runner.jar")  
.withArgs("hbase", "snapshot", "create", "-n", "snapshotName", "-t", "tableName");
```

Note

Si votre nom d'instantané n'est pas unique, l'opération de création échoue avec un code de retour de -1 ou 255, mais vous ne verrez peut-être pas un message d'erreur indiquant ce qui s'est passé. Pour utiliser le même nom d'instantané, supprimez-le et recréez-le.

Suppression d'un instantané

```
hbase shell  
>> delete_snapshot 'snapshotName'
```

Affichage des informations d'instantané

```
hbase snapshot info -snapshot snapshotName
```

Exportation d'un instantané vers Amazon S3

Important

Si vous ne spécifiez aucune `-mappers` valeur lors de l'exportation d'un instantané, HBase utilise un calcul arbitraire pour déterminer le nombre de mappers. Cette valeur peut être très grande, selon la taille de votre table, ce qui affecte de façon négative des tâches exécutées selon l'exportation. Pour cette raison, nous vous recommandons de spécifier le paramètre `-mappers`, le paramètre `-bandwidth` (qui spécifie la consommation de bande passante en mégaoctets par seconde), ou les deux, pour limiter les ressources de cluster utilisées par

l'opération d'exportation. Vous pouvez aussi exécuter l'opération d'instantané d'exportation durant une période de faible utilisation.

```
hbase snapshot export -snapshot snapshotName \
-copy-to s3://bucketName/folder -mappers 2
```

En utilisant `command-runner.jar` depuis AWS CLI :

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar",\
Args=[ "hbase", "snapshot", "export", "-snapshot", "snapshotName", "-copy-to", "s3://bucketName/folder", "-mappers", "2", "-bandwidth", "50"]
```

AWS SDK for Java:

```
HadoopJarStepConfig hbaseImportSnapshotConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("hbase", "snapshot", "export",
        "-snapshot", "snapshotName", "-copy-to",
        "s3://bucketName/folder",
        "-mappers", "2", "-bandwidth", "50");
```

Importer un instantané depuis Amazon S3

Bien qu'il s'agisse d'une importation, l'HBBaseoption utilisée ici l'est toujours `export`.

```
sudo -u hbase hbase snapshot export \
-D hbase.rootdir=s3://bucketName/folder \
-snapshot snapshotName \
-copy-to hdfs://masterPublicDNSName:8020/user/hbase \
-mappers 2
```

En utilisant `command-runner.jar` depuis AWS CLI :

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Name="HBase Shell Step",Jar="command-runner.jar", \
Args=["sudo", "-u", "hbase", "hbase snapshot export", "-snapshot", "snapshotName", \
"-D", "hbase.rootdir=s3://bucketName/folder", \
"-copy-to", "hdfs://masterPublicDNSName:8020/user/hbase", "-mappers", "2", "-chmod", "700"]
```


AWS SDK for Java:

```
HadoopJarStepConfig hbaseImportSnapshotConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("sudo", "-u", "hbase", "hbase", "snapshot", "export", "-D", "hbase.rootdir=s3://
path/to/snapshot",
    "-snapshot", "snapshotName", "-copy-to",
    "hdfs://masterPublicDNSName:8020/user/hbase",
    "-mappers", "2", "-chuser", "hbase");
```

Restaurer une table à partir d'instantanés dans le shell HBase

```
hbase shell
>> disable tableName
>> restore_snapshot snapshotName
>> enable tableName
```

HBase ne prend actuellement pas en charge toutes les commandes de capture d'écran présentes dans le HBase shell. Par exemple, aucune option de HBase ligne de commande ne permet de restaurer un instantané. Vous devez donc le restaurer dans un shell. Cela signifie que `command-runner.jar` doit exécuter une commande Bash.

 Note

Comme la commande utilisée ici l'estecho, il est possible que votre commande shell échoue même si la commande exécutée par Amazon EMR renvoie un code de 0 sortie. Vérifiez les journaux d'étape si vous choisissez d'exécuter une commande shell en tant qu'étape.

```
echo 'disable tableName; \
restore_snapshot snapshotName; \
enable tableName' | hbase shell
```

Voici l'étape utilisant l' AWS CLI. Commencez par créer le fichier `snapshot.json` suivant :

```
[
  {
    "Name": "restore",
    "Args": ["bash", "-c", "echo '$'disable \"tableName\"; restore_snapshot
\"snapshotName\"; enable \"tableName\"] | hbase shell"],
```

```
"Jar": "command-runner.jar",
"ActionOnFailure": "CONTINUE",
"Type": "CUSTOM_JAR"
}
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./snapshot.json
```

AWS SDK for Java:

```
HadoopJarStepConfig hbaseRestoreSnapshotConf = new HadoopJarStepConfig()
.withJar("command-runner.jar")
.withArgs("bash","-c","echo '$disable \"tableName\"; restore_snapshot \"snapshotName\"; enable \"snapshotName\"' | hbase shell");
```

Configurez HBase

Bien que les HBase paramètres par défaut devraient fonctionner pour la plupart des applications, vous pouvez modifier vos paramètres HBase de configuration. Pour ce faire, utilisez les propriétés des classifications de HBase configuration. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

L'exemple suivant crée un cluster avec un répertoire HBase racine alternatif basé sur un fichier de configuration stocké dans Amazon S3. `myConfig.json`

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=HBase \
--instance-type m5.xlarge --instance-count 3 --configurations https://s3.amazonaws.com/
mybucket/myfolder/myConfig.json
```

Le `myConfig.json` fichier spécifie la `hbase.rootdir` propriété de la classification de configuration `hbase-site`, comme illustré dans l'exemple suivant. Remplacez `ip-XXX-XX-XX-XXX.ec2.internal` avec le DNS nom d'hôte interne du nœud principal du cluster.

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
      "hbase.rootdir": "hdfs://ip-XXX-XX-XX-XXX.ec2.internal:8020/user/
myCustomHBaseDir"
    }
  }
]
```

Note

Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier des classifications de configuration supplémentaires pour chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, utilisez la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).

Modifications apportées à l'allocation de mémoire dans YARN

HBase ne s'exécute pas en tant qu'YARN application, il est donc nécessaire de recalculer la mémoire allouée YARN et ses applications, ce qui entraîne une réduction de la mémoire globale disponible pour YARN l'installation. HBase Vous devez en tenir compte lorsque vous planifiez de colocaliser YARN des applications HBase sur les mêmes clusters. Les types d'instance dotés de moins de 64 Go de mémoire disposent de la moitié de la mémoire disponible NodeManager, qui est ensuite allouée au HBase RegionServer. Pour les types d'instance dotés d'une mémoire supérieure à 64 Go, HBase RegionServer la mémoire est limitée à 32 Go. En règle générale, le YARN réglage de la mémoire est un multiple de la mémoire des tâches du MapReduce réducteur.

Les tableaux ci-dessous [Valeurs par défaut des paramètres de configuration des tâches](#) indiquent les modifications apportées aux YARN paramètres en fonction de la mémoire requise pour HBase.

HBasenuméros de port

Certains numéros de port choisis HBase sont différents de ceux par défaut. Les interfaces et les ports suivants sont destinés HBase à AmazonEMR.

HBaseports

utilisateur	Port	Protocole
HMaster	16000	TCP
HMasterUI	16010	HTTP
RegionServer	16020	TCP
RegionServer Infos	16030	HTTP
RESTserveur	8070	HTTP
RESTUI	8085	HTTP
Serveur Thrift	9090	TCP
Interface utilisateur du serveur Thrift	9095	HTTP

Important

`kms-http-port` s'agit du 9700 et du `kms-admin-port` 9701 dans les EMR versions 4.6.0 et ultérieures d'Amazon.

HBaseparamètres du site pour optimiser

Vous pouvez définir tout ou partie des paramètres du HBase site afin d'optimiser le HBase cluster en fonction de la charge de travail de votre application. Nous recommandons les paramètres suivants comme point de départ dans votre enquête.

zookeeper.session.timeout

Le délai d'attente par défaut est de 40 secondes (40 000 ms). En cas de défaillance d'un serveur de région, il s'agit du temps nécessaire au serveur maître pour remarquer l'absence du serveur de région et commencer la récupération. Pour aider le serveur maître à récupérer plus rapidement, vous pouvez réduire cette valeur à une période de temps plus courte. L'exemple suivant utilise 30 secondes ou 30 000 ms :

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
      "zookeeper.session.timeout": "30000"
    }
  }
]
```

hbase.regionserver.handler.count

Cela définit le nombre de threads que le serveur de région laisse ouverts pour diffuser les demandes aux tables. La valeur par défaut de 10 est faible, afin d'empêcher les utilisateurs de supprimer leurs serveurs de région lorsqu'ils utilisent de grands tampons d'écriture avec un grand nombre de clients simultanés. La règle de base est de maintenir ce chiffre à un niveau bas lorsque la charge utile par demande se rapproche de l'ordre de Mo (gros achats, scans utilisant un cache volumineux) et élevé lorsque la charge utile est faible (get, small put/ICVs, delete). L'exemple suivant augmente le nombre de threads ouverts à 30 :

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
      "hbase.regionserver.handler.count": "30"
    }
  }
]
```

hbase.hregion.max.filesize

Ce paramètre régit la taille, en octets, des régions individuelles. Par défaut, l'attribut est défini sur 1073741824. Si vous écrivez beaucoup de données dans votre HBase cluster et que cela entraîne

des fractionnements fréquents, vous pouvez augmenter cette taille pour agrandir les différentes régions. Cela réduit les divisions mais augmente le temps nécessaire pour équilibrer la charge des régions d'un serveur à un autre.

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
      "hbase.hregion.max.filesize": "1073741824"
    }
  }
]
```

hbase.hregion.memstore.flush.size

Ce paramètre régit la taille maximale de memstore, en octets, avant son vidage sur le disque. Par défaut, il s'agit de 134217728. Si votre charge de travail se compose de courtes périodes intensives d'opérations d'écriture, vous pouvez augmenter cette limite afin que toutes les écritures restent en mémoire au cours de ces périodes et soient vidées sur le disque plus tard. Cela peut améliorer les performances au cours de ces périodes intensives.

```
[
  {
    "Classification":"hbase-site",
    "Properties": {
      "hbase.hregion.memstore.flush.size": "134217728"
    }
  }
]
```

Afficher l'interface HBase utilisateur

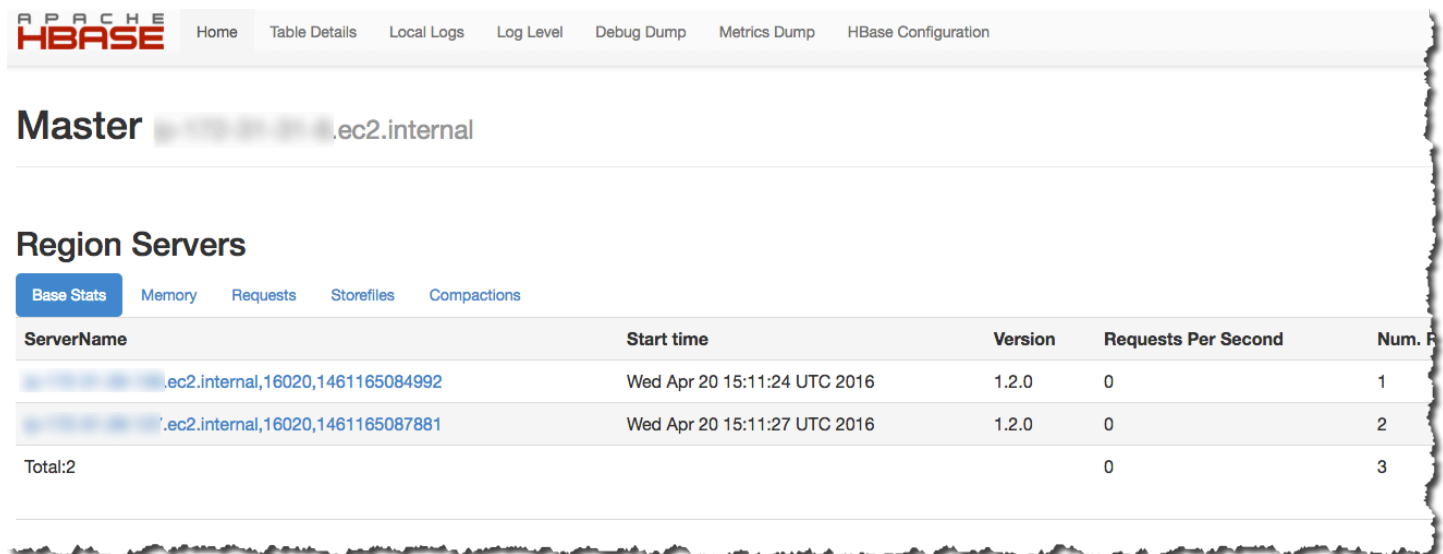
Note

L'interface HBase utilisateur utilise des HTTP connexions non sécurisées par défaut. Pour activer secure HTTP (HTTPS), définissez la `hbase.ssl.enabled` propriété de la `hbase-site` classification sur `true` dans votre [HBaseconfiguration](#). Pour plus d'informations sur l'utilisation de secure HTTP (HTTPS) pour l'interface utilisateur HBase Web, consultez le [guide de HBase référence Apache](#).

HBase fournit une interface utilisateur Web que vous pouvez utiliser pour surveiller votre HBase cluster. Lorsque vous utilisez AmazonEMR, l'interface Web s'exécute sur le nœud principal et peut être consultée à l'aide de la redirection de port, également appelée création d'un SSH tunnel. HBase

Pour afficher l'interface HBase utilisateur

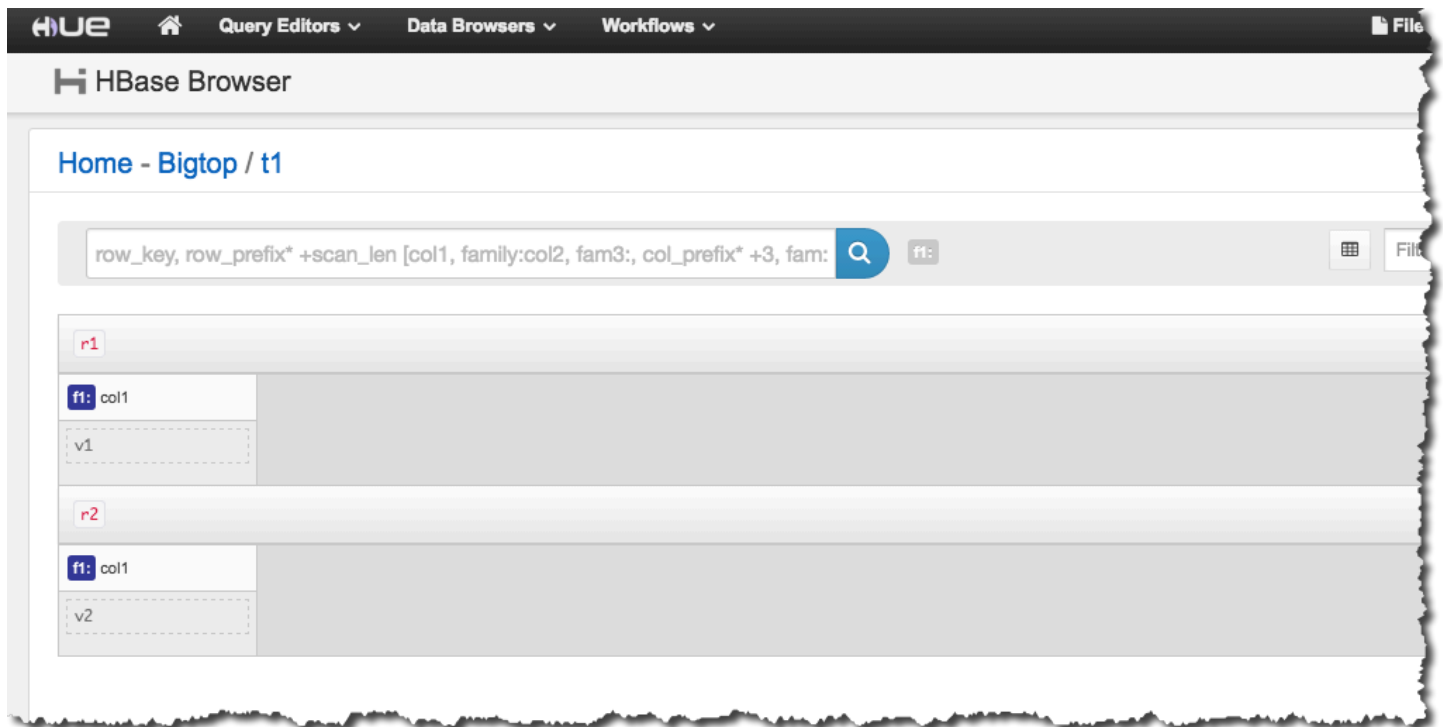
1. SSH à utiliser pour créer un tunnel vers le nœud principal et créer une connexion sécurisée. Pour plus d'informations, consultez [l'option 2, partie 1 : configurer un SSH tunnel vers le nœud principal à l'aide de la redirection de port dynamique](#) dans le guide EMR de gestion Amazon.
2. Installez un navigateur Web avec un outil proxy, tel que le FoxyProxy plug-in pour Firefox, afin de créer un SOCKS proxy pour les AWS domaines. Pour plus d'informations, consultez [l'option 2, partie 2 : configurer les paramètres du proxy pour afficher les sites Web hébergés sur le nœud principal](#) dans le guide EMR de gestion Amazon.
3. Une fois le proxy défini et la SSH connexion ouverte, vous pouvez afficher l'HBase interface utilisateur en ouvrant une fenêtre de navigateur avec `http://master-public-dns-name:16010/master-status`, où *master-public-dns-name* est l'adresse DNS publique du nœud principal du cluster.



The screenshot shows the Apache HBase web interface. At the top, there is a navigation bar with links: Home, Table Details, Local Logs, Log Level, Debug Dump, Metrics Dump, and HBase Configuration. Below the navigation bar, the page title is "Master" followed by a blurred IP address and ".ec2.internal". The main content area is titled "Region Servers" and has several tabs: Base Stats (selected), Memory, Requests, Storefiles, and Compactions. Below the tabs is a table with the following data:

ServerName	Start time	Version	Requests Per Second	Num. F
<i>blurred</i> .ec2.internal,16020,1461165084992	Wed Apr 20 15:11:24 UTC 2016	1.2.0	0	1
<i>blurred</i> .ec2.internal,16020,1461165087881	Wed Apr 20 15:11:27 UTC 2016	1.2.0	0	2
Total:2			0	3

Vous pouvez également le voir HBase à Hue. Par exemple, l'illustration suivante montre la table, t1, que vous avez créée dans [Utilisation de la HBase coque](#) :



Pour plus d'informations sur Hue, consultez [Hue](#).

Afficher les fichiers journaux HBase

Dans le cadre de son fonctionnement, HBase écrit des fichiers journaux contenant des détails sur les paramètres de configuration, les actions du démon et les exceptions. Ces fichiers journaux peuvent être utiles pour le débogage des problèmes HBase ainsi que pour le suivi des performances.

Si vous configurez votre cluster pour conserver les fichiers journaux dans Amazon S3, vous devez savoir que les journaux sont écrits dans Amazon S3 toutes les cinq minutes, si bien qu'il peut y avoir un léger retard avant que les derniers fichiers journaux soient disponibles.

Pour afficher HBase les journaux sur le nœud principal

- Vous pouvez consulter les HBase journaux actuels en vous connectant SSH au nœud principal et en accédant au `/var/log/hbase` répertoire. Ces journaux ne sont pas disponibles une fois que le cluster a été arrêté sauf si vous activez la journalisation dans Amazon S3 lorsque le cluster est lancé.

Pour consulter HBase les journaux sur Amazon S3

- Pour accéder aux HBase journaux et aux autres journaux de cluster sur Amazon S3, et pour qu'ils soient disponibles après la fin du cluster, spécifiez un compartiment Amazon S3 pour recevoir ces journaux lorsque vous créez le cluster. Pour cela, utilisez l'option `--log-uri`. Pour plus d'informations sur l'activation de la journalisation pour votre cluster, consultez [Configurer la journalisation et le débogage \(facultatif\)](#) dans le guide EMR de gestion Amazon.

Moniteur HBase avec Ganglia

Le projet open source Ganglia est un système distribué évolutif, conçu pour surveiller les clusters et les grilles tout en réduisant au maximum l'impact sur leurs performances. Lorsque vous activez Ganglia sur votre cluster, vous pouvez générer des rapports et afficher la performance du cluster dans son ensemble, ainsi qu'inspecter la performance des instances de chaque nœud. Pour plus d'informations sur le projet open source Ganglia, consultez <http://ganglia.info/>. Pour plus d'informations sur l'utilisation de Ganglia avec les EMR clusters Amazon, consultez [Ganglia](#).

Une fois le cluster lancé avec Ganglia configuré, vous pouvez accéder aux graphes et aux rapports Ganglia à l'aide de l'interface graphique exécutée sur le nœud principal.

Ganglia stocke les fichiers journaux sur le nœud principal du `/mnt/var/lib/ganglia/rrds/` répertoire. Les versions antérieures d'Amazon EMR peuvent stocker des fichiers journaux dans le `/var/log/ganglia/rrds/` répertoire.

Pour configurer un cluster pour Ganglia et à HBase l'aide du AWS CLI

- Utilisez une commande `create-cluster` similaire à la commande suivante :

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \  
--applications Name=HBase Name=Ganglia --use-default-roles \  
--ec2-attributes KeyName=myKey --instance-type m5.xlarge \  
--instance-count 3
```

Note

Si le rôle de EMR service Amazon par défaut et le profil d'EC2instance Amazon n'existent pas, une erreur se produit. Utilisez la commande `aws emr create-default-roles` pour les créer, puis réessayez.

Pour plus d'informations, consultez [EMR les commandes Amazon dans le AWS CLI](#).

Pour afficher les HBase métriques dans l'interface Web de Ganglia

1. SSH à utiliser pour créer un tunnel vers le nœud principal et créer une connexion sécurisée. Pour plus d'informations, consultez l'[option 2, partie 1 : configurer un SSH tunnel vers le nœud principal à l'aide de la redirection de port dynamique](#) dans le guide EMR de gestion Amazon.
2. Installez un navigateur Web avec un outil proxy, tel que le FoxyProxy plug-in pour Firefox, afin de créer un SOCKS proxy pour les AWS domaines. Pour plus d'informations, consultez l'[option 2, partie 2 : configurer les paramètres du proxy pour afficher les sites Web hébergés sur le nœud principal](#) dans le guide EMR de gestion Amazon.
3. Une fois le proxy défini et la SSH connexion ouverte, vous pouvez consulter les métriques de Ganglia en ouvrant une fenêtre de navigateur avec `http ://master-public-dns-name/ganglia/`, où *master-public-dns-name* est l'adresse DNS publique du serveur principal du HBase cluster.

Pour afficher les fichiers journaux Ganglia sur le nœud principal

- Si le cluster est toujours en cours d'exécution, vous pouvez accéder aux fichiers journaux en SSH vous connectant au nœud principal et en accédant au `/mnt/var/lib/ganglia/rrds/` répertoire. Pour EMR 3.x, accédez au `/var/log/ganglia/rrds` répertoire. Pour plus d'informations, consultez [Se connecter au nœud principal SSH à l'aide](#) de l'Amazon EMR Management Guide.

Pour afficher les fichiers journaux de Ganglia sur Amazon S3

- Les fichiers journaux Ganglia ne sont pas automatiquement écrits dans Amazon S3 même si vous activez la journalisation pour votre cluster. Pour afficher les fichiers journaux Ganglia sur Amazon S3, vous devez les pousser manuellement de `/mnt/var/lib/ganglia/rrds/` vers le compartiment S3.

Migration depuis les versions précédentes HBase

Pour migrer les données d'une HBase version précédente, consultez la section [Mise à niveau, numéro de HBase version et compatibilité](#) dans le guide de HBase référence Apache. Vous devez peut-être porter une attention particulière aux exigences relatives à la mise à niveau à partir des versions antérieures à 1.0 de HBase.

HBase historique des versions

Le tableau suivant répertorie la version HBase incluse dans chaque version d'Amazon EMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

HBase informations sur la version

Étiquette Amazon EMR Release	HBase Version	Composants installés avec HBase
emr-7.2.0	2.4,17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-ma pred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn-timeline-server, hbase-hma ster, hbase-client, hbase-reg ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope rator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.36.2	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-7.1.0	2.4,17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, emr-wal-cli, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-7.0.0	2.4,17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma pred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn-timeline-server, hbase-hma ster, hbase-client, hbase-reg ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope rator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.15.0	2.4,17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.14.0	2.4,17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma pred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn-timeline-server, hbase-hma ster, hbase-client, hbase-reg ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope rator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.13.0	2.4,17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.12.0	2.4,17	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma pred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn-timeline-server, hbase-hma ster, hbase-client, hbase-reg ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope rator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.11.1	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.11.0	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma pred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn-timeline-server, hbase-hma ster, hbase-client, hbase-reg ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope rator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.10.1	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.10.0	2.4.15	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-wal-cli, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-ma pred, hadoop-yarn-nodema nager, hadoop-yarn-resour cemanager, hadoop-yarn-timeline-server, hbase-hma ster, hbase-client, hbase-reg ion-server, hbase-rest-server, hbase-thrift-server, hbase-ope rator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.9.1	2.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-6.9.0	2.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.8.1	2.4.12	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-6.8.0	2.4.12	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.7.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server
emr-5.36.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.36.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.6.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, hbase-operator-tools, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.35.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.5.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.4.0	2.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.3.1	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.3.0	2.2.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.2.1	2.2.6-amzn-0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.2.0	2.2.6-amzn-0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.1.1	2.2.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.1.0	2.2.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-6.0.1	2.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-6.0.0	2.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.34.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.33.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.33.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.32.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.32.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.31.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.31.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.30.2	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.30.1	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.30.0	1.4.13	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
mr-5.29.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.28.1	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.28.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.27.1	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.27.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.26.0	1.4.10	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.25.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.24.1	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.24.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.23.1	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.23.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.22.0	1.4.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.2	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.21.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.21.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.20.1	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.20.0	1.4.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.19.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.19.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.17.2	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.17.1	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.17.0	1.4.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.16.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.16.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.15.1	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.15.0	1.4.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.2	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.14.1	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.0	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.13.1	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.13.0	1.4.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.12.3	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.2	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.12.1	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.0	1.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.11.4	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.11.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.11.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.10.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.10.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.9.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.9.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.3	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.8.2	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.8.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.7.1	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.7.0	1.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.6.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.6.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.4	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.5.3	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.2	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.5.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.5.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.4.1	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.4.0	1.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.3.2	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.3.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.3.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.2.3	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.2.2	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.2.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.2.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.1.1	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.1.0	1.2.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.0.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.0.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.0.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-5.0.0	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.6	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-4.9.5	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.4	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-4.9.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.9.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-4.9.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.5	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-4.8.4	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.3	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-4.8.2	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.8.1	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-4.8.0	1.2.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.4	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-4.7.3	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.2	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-4.7.1	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.7.0	1.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	HBaseVersion	Composants installés avec HBase
emr-4.6.1	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-4.6.0	1.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Apache HCatalog

HCatalog est un outil qui vous permet d'accéder aux tables de métastore Hive dans PigSQL, Spark et/ou dans des applications personnalisées. MapReduce HCatalog possède une REST interface et un client de ligne de commande qui vous permettent de créer des tables ou d'effectuer d'autres opérations. Vous écrivez ensuite vos applications pour accéder aux tables à l'aide de HCatalog bibliothèques. Pour plus d'informations, veuillez consulter [Utilisation de HCatalog](#). HCatalog est inclus dans les EMR versions 4.4.0 et ultérieures d'Amazon.

HCatalog sur Amazon, les EMR versions 5.8.0 et ultérieures prennent en charge l'utilisation de AWS Glue Data Catalog comme métastore pour Hive. Pour plus d'informations, consultez la section [Utilisation du catalogue de données AWS Glue comme métastore pour Hive](#).

Le tableau suivant répertorie la version HCatalog incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants avec lesquels Amazon EMR installe HCatalog.

Pour connaître la version des composants installés HCatalog dans cette version, consultez la [version 6.15.0 Versions des composants](#).

HCatalog informations de version pour emr-6.15.0

Étiquette EMR de sortie Amazon	HCatalog Version	Composants installés avec HCatalog
emr-6.15.0	HCatalog 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Le tableau suivant répertorie la version HCatalog incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants avec lesquels Amazon EMR installe HCatalog

Pour la version des composants installés HCatalog dans cette version, voir la [version 5.36.2 Versions des composants](#).

HCatalog informations de version pour emr-5.36.2

Étiquette EMR de sortie Amazon	HCatalogVersion	Composants installés avec HCatalog
emr-5.36.2	HCatalog 2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Rubriques

- [Création d'un cluster avec HCatalog](#)
- [En utilisant HCatalog](#)
- [Exemple : créer une HCatalog table et y écrire en utilisant Pig](#)
- [HCatalog historique des versions](#)

Création d'un cluster avec HCatalog

Bien qu'HCatalog soit inclus dans le projet Hive, vous devez l'installer en tant qu'application autonome.

Pour lancer un cluster HCatalog installé à l'aide de la console

La procédure suivante crée un cluster avec HCatalog installé. Pour plus d'informations sur la création de clusters à l'aide de la console, notamment sur les options avancées, consultez [Planifier et configurer des clusters](#) dans le guide EMR de gestion Amazon.

1. Ouvrez la EMR console Amazon à l'[adresse https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr).
2. Choisissez Create cluster (Créer un cluster) pour utiliser la Quick Create (Création rapide).
3. Pour le champ Software Configuration (Configuration des logiciels), choisissez la Version Amazon emr-4.4.0 ou une version ultérieure.
4. Dans le champ Sélectionner les applications, sélectionnez Toutes les applications ou HCatalog.
5. Sélectionnez d'autres options si nécessaire, puis choisissez Create cluster (Créer le cluster).

Pour lancer un cluster à HCatalog l'aide du AWS CLI

- Créez le cluster à l'aide de la commande suivante :

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

```
aws emr create-cluster --name "Cluster with Hcat" --release-label emr-7.2.0 \  
--applications Name=HCatalog --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

En utilisant HCatalog

Vous pouvez l'utiliser HCatalog dans différentes applications qui utilisent le métastore Hive. Les exemples de cette section montrent comment créer une table et l'utiliser dans le contexte de Pig et SparkSQL.

Désactiver l'écriture directe lors de l'utilisation HCatalog HStorer

Chaque fois qu'une [HCatStorer](#) application écrit dans une HCatalog table stockée dans Amazon S3, désactivez la fonctionnalité d'écriture directe d'AmazonEMR. Par exemple, désactivez l'écriture directe lorsque vous utilisez la STORE commande Pig ou lorsque vous exécutez des tâches Sqoop qui écrivent HCatalog des tables sur Amazon S3. Vous pouvez désactiver la fonctionnalité d'écriture directe en définissant `mapred.output.direct.NativeS3FileSystem` et les configurations `mapred.output.direct.EmrFileSystem` sur `false`. L'exemple suivant montre comment définir ces configurations à l'aide de Java.

```
Configuration conf = new Configuration();
conf.set("mapred.output.direct.NativeS3FileSystem", "false");
conf.set("mapred.output.direct.EmrFileSystem", "false");
```

Créez une table en utilisant HCat CLI et utilisez ces données dans Pig

Créez le script suivant, `impressions.q`, sur votre cluster :

```
CREATE EXTERNAL TABLE impressions (
  requestBeginTime string, adId string, impressionId string, referrer string,
  userAgent string, userCookie string, ip string
)
PARTITIONED BY (dt string)
ROW FORMAT
  serde 'org.apache.hive.hcatalog.data.JsonSerDe'
  with serdeproperties ( 'paths'='requestBeginTime, adId, impressionId, referrer,
userAgent, userCookie, ip' )
LOCATION 's3://[your region].elasticmapreduce/samples/hive-ads/tables/impressions/';
ALTER TABLE impressions ADD PARTITION (dt='2009-04-13-08-05');
```

Exécutez le script en utilisant HCat CLI :

```
% hcat -f impressions.q
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-
log4j.properties
OK
Time taken: 4.001 seconds
OK
Time taken: 0.519 seconds
```

Ouvrez le shell Grunt et accédez aux données dans `impressions` :

```
% pig -useHCatalog -e "A = LOAD 'impressions' USING
  org.apache.hive.hcatalog.pig.HCatLoader();
B = LIMIT A 5;
dump B;"
<snip>
(1239610346000,m9nwdo67Nx6q2kI25qt50n7peICfUM,omkxkaRpNhGPDucAiBErSh1cs0MThC,cartoonnetwork.com
  (compatible; MSIE 7.0; Windows NT 6.0; FunWebProducts; GTB6; SLCC1; .NET CLR
  2.0.50727; Media Center PC
  5.0; .NET,wcVWWTascoPbGt6bdqDbuWTPPHgOPs,69.191.224.234,2009-04-13-08-05)
(1239611000000,NjriQjd0DgWBKnkGJUP6GNTbDeK4An,AwtxPkfaWG0aNeL900sFU8Hcj6eLHT,cartoonnetwork.com
  (compatible; MSIE 7.0; Windows NT 5.1; GTB6; .NET CLR
  1.1.4322),0aMU1F2gE4CtADVHAbKjjRRks5kIgg,57.34.133.110,2009-04-13-08-05)
(1239610462000,Irpv3oiu0I5QNQiWSSIIshrLdo9cM1,i1LDq44LRSJF0hbmhB8Gk7k9gMWtBq,cartoonnetwork.com
  (compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
  InfoPath.1),Qsb3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239611007000,q2Awfnpe0JAvhInaIp0VGx9Kts0oPO,s3HvTf1PB8JIE0IuM6h0EebWwP0tJV,cartoonnetwork.com
  (compatible; MSIE 6.0; Windows NT 5.2; SV1; .NET CLR 1.1.4322;
  InfoPath.1),Qsb3wkLR4JAIut4Uq6FNFQIR1rCVwU,42.174.193.253,2009-04-13-08-05)
(1239610398000,c362vpAB0soPKGHRs43cj6TRwNe0Gn,jeas5nXbQInGAgFB8jlkhnprN6cMw7,cartoonnetwork.com
  (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0; GTB6; .NET CLR
  1.1.4322),k96n5PnUmwHKfiUI0TFP0TNMFADgh9,51.131.29.87,2009-04-13-08-05)
7120 [main] INFO org.apache.pig.Main - Pig script completed in 7 seconds and 199
  milliseconds (7199 ms)
16/03/08 23:17:10 INFO pig.Main: Pig script completed in 7 seconds and 199 milliseconds
  (7199 ms)
```

Accès à la table à l'aide de Spark SQL

Cet exemple crée un Spark DataFrame à partir de la table créée dans le premier exemple et montre les 20 premières lignes :

```
% spark-shell --jars /usr/lib/hive-hcatalog/share/hcatalog/hive-hcatalog-core-1.0.0-
  amzn-3.jar
<snip>
scala> val hiveContext = new org.apache.spark.sql.hive.HiveContext(sc);
scala> val df = hiveContext.sql("SELECT * FROM impressions")
scala> df.show()
<snip>
16/03/09 17:18:46 INFO DAGScheduler: ResultStage 0 (show at <console>:32) finished in
  10.702 s
16/03/09 17:18:46 INFO DAGScheduler: Job 0 finished: show at <console>:32, took
  10.839905 s
```



```

+-----+-----+-----+-----+
+-----+-----+-----+-----+
|requestbegintime|          adid|          impressionid|          referrer|
|  useragent|          usercookie|          ip|          dt|
+-----+-----+-----+-----+
+-----+-----+-----+-----+
|  1239610346000|m9nwd067Nx6q2kI25...|omkxkaRpNhGPDucAi...|cartoonnetwork.com|
Mozilla/4.0 (comp...|wcVWWTascoPbGt6bd...|69.191.224.234|2009-04-13-08-05|
|  1239611000000|NjriQjd0DgWBKnkGJ...|AWtXPKfawG0aNeL90...|cartoonnetwork.com|
Mozilla/4.0 (comp...|0aMU1F2gE4CtADVHA...| 57.34.133.110|2009-04-13-08-05|
|  1239610462000|Irpv3oiu0I5QNQiW...|i1LDq44LRSJF0hbmh...|cartoonnetwork.com|
Mozilla/4.0 (comp...|QSb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
|  1239611007000|q2Awfnpe0JAvhInaI...|s3HvTf1PB8JIE0IuM...|cartoonnetwork.com|
Mozilla/4.0 (comp...|QSb3wkLR4JAIut4Uq...|42.174.193.253|2009-04-13-08-05|
|  1239610398000|c362vpAB0soPKGHR...|jeas5nXbQInGAgFB8...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610600000|cjbTpruoaiEtqLuMX...|XwlohBSs8Ipxs1bRa...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610804000|Ms3eJHNAEItpxvimd...|4SIj4pGmgVL1625BD...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610872000|h5bccHX6wJReDi1jL...|EFAWIiBdVfnxwAMWP...|cartoonnetwork.com|
Mozilla/4.0 (comp...|k96n5PnUmwHKfiUI0...| 51.131.29.87|2009-04-13-08-05|
|  1239610365000|874NBpGmxNFfxEPKM...|xSvE4XtGbdTXPF2Lb...|cartoonnetwork.com|
Mozilla/5.0 (Maci...|eWDEVVUphlnRa273j...| 22.91.173.232|2009-04-13-08-05|
|  1239610348000|X8gISpUTSqh1A5reS...|TrFblGT99AgE75vuj...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610743000|kbKreLWB6QVueFrDm...|kVnxx9Ie2i30LTxFj...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610812000|9lX0SRpEi3bmEeTCu...|1B2sff99AEIwSuLVV...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610876000|lijjmCf2kuxfBTnjL...|AjvufgUtakUFcsIM9...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610941000|t8t8trgjNRPIlmxuD...|agu2u2TCdqWP08rAA...|          corriere.it|
Mozilla/4.0 (comp...|tX1sMpnhJUhmAF7AS...| 55.35.44.79|2009-04-13-08-05|
|  1239610490000|OGRLPVNGxiGgrCmWL...|mJg2raBUpPrC80lUm...|          corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...| 71.124.66.3|2009-04-13-08-05|
|  1239610556000|OnJID12x0RXKPUgrD...|P7Pm2mPdW6w08KA3R...|          corriere.it|
Mozilla/4.0 (comp...|r2k96t1CNjSU9fJKN...| 71.124.66.3|2009-04-13-08-05|
|  1239610373000|WflsvKIg0qfIE5KwR...|TJHd1VBspNcua0XPn...|          corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
|  1239610768000|4MJR0XxiVCU1ueXKV...|10hGwmbvKf8ajoU8a...|          corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|
|  1239610832000|gWIrpDiN57i3sHatv...|RNL4C7xPi3tdar2Uc...|          corriere.it|
Mozilla/5.0 (Maci...|fj2L1ILTFGMfhdrt3...| 75.117.56.155|2009-04-13-08-05|

```

```
| 1239610789000|pTne9k62kJ14QViXI...|RVxJVIQousjxUVI3r...| pixnet.net|
Mozilla/5.0 (Maci...|1bG0KiBD2xmui90kF...| 33.176.101.80|2009-04-13-08-05|
+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+
only showing top 20 rows

scala>
```

Exemple : créer une HCatalog table et y écrire en utilisant Pig

Vous pouvez créer une HCatalog table et utiliser Apache Pig pour y écrire en HCatStorer utilisant une source de données dans Amazon S3. HCatalognécessite que vous désactiviez l'écriture directe, sinon l'opération échoue silencieusement. Définissez les configurations `mapred.output.direct.NativeS3FileSystem` et `mapred.output.direct.EmrFileSystem` sur `false` à l'aide de la classification `mapred-site`, ou manuellement depuis le shell Grunt. L'exemple suivant montre une table créée à l'aide de HCatCLI, suivie de commandes exécutées dans le shell Grunt pour remplir la table à partir d'un exemple de fichier de données dans Amazon S3.

Pour exécuter cet exemple, [connectez-vous au nœud principal à l'aide deSSH](#).

Créez un fichier de HCatalog script `wikicount.q`, avec le contenu suivant, qui crée une HCatalog table nommée `wikicount`.

```
CREATE EXTERNAL TABLE IF NOT EXISTS wikicount(
  col1 string,
  col2 bigint
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\001'
STORED AS ORC
LOCATION 's3://MyBucket/hcat/wikicount';
```

Utilisez une HCat CLI commande pour exécuter le script à partir du fichier.

```
hcat -f wikicount.q
```

Ensuite, démarrez le shell Grunt avec l'option `-useHCatalog`, définissez des configurations pour désactiver l'écriture directe, chargez les données depuis un emplacement S3, puis écrivez les résultats dans la table `wikicount`.

```
pig -useHCatalog
SET mapred.output.direct.NativeS3FileSystem false;
SET mapred.output.direct.EmrFileSystem false;
A = LOAD 's3://support.elasticmapreduce/training/datasets/wikistats_tiny/' USING
  PigStorage(' ') AS (Site:chararray, page:chararray, views:int, total_bytes:long);
B = GROUP A BY Site;
C = FOREACH B GENERATE group as col1, COUNT(A) as col2;
STORE C INTO 'wikicount' USING org.apache.hive.hcatalog.pig.HCatStorer();
```

HCataloghistorique des versions

Le tableau suivant répertorie la version HCatalog incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section [Version des composants de votre version](#) dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

HCataloginformations sur la version

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-7.2.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-s

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
		erver, hive-client, mariadb-server
emr-5.36.2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-7.1.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-7.0.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.15.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.13.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.12.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.11.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.11.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.10.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.10.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.9.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.9.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.8.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.8.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.7.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.36.1	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.36.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.6.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.35.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.5.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.4.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.3.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.3.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.2.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.2.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.1.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.1.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-6.0.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-6.0.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.34.0	2.3.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.33.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.32.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.31.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.30.2	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server
emr-5.30.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mariadb-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
mr-5.29.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.28.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.27.1	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.24.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.23.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.22.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.21.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.20.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.19.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.18.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.17.2	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.17.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.16.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.16.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.15.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.15.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.14.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.14.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.14.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.13.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.13.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.12.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.12.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.12.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.11.4	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.11.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.11.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.11.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.11.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.10.1	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.10.0	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.9.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.9.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.8.3	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.8.2	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.8.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.8.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.7.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.7.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.6.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.6.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.5.4	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.5.3	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.5.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.5.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.4.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.4.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.3.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.3.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.3.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.2.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.2.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.2.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.2.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.1.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.0.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.0.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-5.0.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-5.0.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.9.6	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.9.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.9.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.9.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.8.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.8.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.8.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.7.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.7.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.7.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, mysql-server
emr-4.6.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Étiquette Amazon EMR Release	HCatalogVersion	Composants installés avec HCatalog
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hcatalog-client, hcatalog-server, hcatalog-webhcat-server, hive-client, hive-metastore-server, mysql-server

Apache Hive

Hive est un entrepôt de données open source et un package analytique qui s'exécute au-dessus d'un cluster Hadoop. Les scripts Hive utilisent un langage SQL similaire appelé Hive QL (langage de requête) qui fait abstraction des modèles de programmation et prend en charge les interactions typiques des entrepôts de données. Hive vous permet d'éviter les difficultés liées à l'écriture de jobs Tez basés sur des graphes acycliques dirigés (DAGs) ou des MapReduce programmes dans un langage informatique de niveau inférieur, tel que Java.

Hive étend le SQL paradigme en incluant des formats de sérialisation. Vous pouvez également personnaliser le traitement d'une requête en créant un schéma de table correspondant à vos données, sans toucher au données elles-même. Bien SQL qu'elles ne prennent en charge que les types de valeurs primitifs (tels que les dates, les nombres et les chaînes), les valeurs des tables Hive sont des éléments structurés, tels que JSON des objets, tout type de données défini par l'utilisateur ou toute fonction écrite en Java.

Pour plus d'informations sur Hive, consultez le site <http://hive.apache.org/>.

Le tableau suivant répertorie la version de Hive incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Hive.

Pour la version des composants installés avec Hive dans cette version, voir [Versions des composants de la version 7.2.0](#).

Informations sur la version de Hive pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version de Hive	Composants installés avec Hive
emr-7.2.0	Hive 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-ya

Étiquette EMR de sortie Amazon	Version de Hive	Composants installés avec Hive
		rn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Le tableau suivant répertorie la version de Hive incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Hive.

Pour connaître la version des composants installés avec Hive dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Hive pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version de Hive	Composants installés avec Hive
emr-6.15.0	Hive 3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Le tableau suivant répertorie la version de Hive incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Hive.

Pour la version des composants installés avec Hive dans cette version, voir la [version 5.36.2](#)
Versions des composants.

Informations sur la version de Hive pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version de Hive	Composants installés avec Hive
emr-5.36.2	Hive 2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

À partir d'Amazon EMR 5.18.0, vous pouvez utiliser le référentiel d'EMRartefacts Amazon pour créer votre code de travail en fonction des versions exactes des bibliothèques et des dépendances disponibles avec des versions spécifiques d'Amazon. EMR Pour plus d'informations, consultez [Vérification des dépendances à l'aide du référentiel d'artefacts d'Amazon EMR](#).

Rubriques

- [Différences et considérations relatives à Hive sur Amazon EMR](#)
- [Configuration d'un metastore externe pour Hive](#)
- [Utilisation du pilote Hive JDBC](#)
- [Amélioration des performances de Hive](#)
- [Utilisation de Hive Live Long et Process \(\) LLAP](#)

- [Chiffrement dans Hive](#)
- [Historique des versions de Hive](#)

Différences et considérations relatives à Hive sur Amazon EMR

Différences entre Apache Hive sur Amazon EMR et Apache Hive

[Cette section décrit les différences entre Hive on Amazon EMR et les versions par défaut de Hive disponibles sur <http://svn.apache.org/viewvc/hive/branches/>.](#)

Autorisation Hive

Amazon EMR prend en charge [l'autorisation Hive](#) pour Amazon S3HDFS, mais pas pour EMRFS Amazon S3. EMR Les clusters Amazon s'exécutent avec l'autorisation désactivée par défaut.

Comportement de fusion de fichiers Hive avec Amazon S3

Apache Hive fusionne les petits fichiers à la fin d'une tâche de type map-only si `hive.merge.mapfiles` est défini sur `true` et la fusion est déclenchée uniquement si la taille de sortie moyenne de la tâche est inférieure au paramètre `hive.merge.smallfiles.avgsize`. Amazon EMR Hive a exactement le même comportement si le chemin de sortie final est entré. HDFS Si le chemin de sortie se trouve dans Amazon S3, le paramètre `hive.merge.smallfiles.avgsize` est ignoré. Dans ce cas, la tâche de fusion est toujours déclenchée si `hive.merge.mapfiles` est défini sur `true`.

ACID transactions et Amazon S3

Amazon EMR 6.1.0 et versions ultérieures prennent en charge les transactions Hive ACID (Atomicity, Consistency, Isolation, Durability) afin de respecter les ACID propriétés d'une base de données. Grâce à cette fonctionnalité, vous pouvez exécuter INSERT, UPDATEDELETE, et MERGE des opérations dans des tables gérées par Hive avec des données dans Amazon Simple Storage Service (Amazon S3).

Hive Live Long et Process () LLAP

[LLAPles fonctionnalités](#) ajoutées dans la version 2.0 d'Apache Hive par défaut ne sont pas prises en charge dans Hive 2.1.0 sur Amazon EMR version 5.0.

Amazon EMR version 6.0.0 et versions ultérieures prend en charge les fonctionnalités Live Long et Process (LLAP) pour Hive. Pour plus d'informations, consultez la section [Utilisation de Hive. LLAP](#)

Différences dans Hive entre les EMR versions 4.x et 5.x d'Amazon

Cette section décrit les différences à prendre en compte avant de migrer une implémentation Hive de la version 1.0.0 de Hive sur Amazon EMR version 4.x vers Hive 2.x sur Amazon version 5.x. EMR

Différences et considérations opérationnelles

- Support ajouté pour les [transactions ACID \(atomicité, cohérence, isolation et durabilité\)](#) : cette différence entre Hive 1.0.0 sur Amazon EMR 4.x et Apache Hive par défaut a été supprimée.
- Élimination des écritures directes sur Amazon S3 : cette différence entre Hive 1.0.0 sur Amazon EMR et Apache Hive par défaut a été supprimée. Hive 2.1.0 sur Amazon EMR version 5.x crée, lit et écrit désormais dans des fichiers temporaires stockés dans Amazon S3. Par conséquent, pour lire et écrire dans la même table, il n'est plus nécessaire de créer une table temporaire dans le système de HDFS fichiers local du cluster pour contourner le problème. Si vous utilisez des compartiments activés pour la gestion des versions, veillez à gérer ces fichiers temporaires comme indiqué ci-dessous.
- Gestion des fichiers temporaires lors de l'utilisation de compartiments Amazon S3 versionnés : lorsque vous exécutez des requêtes Hive pour lesquelles la destination des données générées est Amazon S3, de nombreux fichiers temporaires et répertoires sont créés. Il s'agit du nouveau comportement, comme indiqué précédemment. Si vous utilisez des compartiments S3 activés pour la gestion des versions, ces fichiers temporaires encombrement Amazon S3 et entraînent des coûts s'ils ne sont pas supprimés. Ajustez vos règles de cycle de vie afin que les données ayant un préfixe `/_tmp` soient supprimées après un délai court, par exemple cinq jours. Pour plus d'informations, consultez [Configuration de cycle de vie](#).
- Mise à jour de Log4j par log4j 2 : si vous utilisez log4j, vous devrez peut-être modifier votre configuration de journalisation en raison de cette mise à niveau. Pour plus d'informations, consultez [Apache log4j 2](#).

Différences et considérations en termes de performances

- Différences de performances avec Tez : avec la EMR version 5.x d'Amazon, Tez est le moteur d'exécution par défaut pour Hive au lieu de. MapReduce Tez offre des performances améliorées pour la plupart des flux de travail.
- Tables avec plusieurs partitions : les requêtes qui génèrent un grand nombre de partitions dynamiques peuvent échouer et les requêtes qui effectuent des sélections sur des tables comportant un grand nombre de partitions peuvent nécessiter des temps d'exécution plus longs

que prévu. Par exemple, une sélection effectuée sur 100 000 partitions peut prendre 10 minutes ou plus.

Fonctionnalités supplémentaires de Hive sur Amazon EMR

Amazon EMR étend Hive avec de nouvelles fonctionnalités qui prennent en charge l'intégration de Hive à d'autres AWS services, telles que la possibilité de lire et d'écrire sur Amazon Simple Storage Service (Amazon S3) et DynamoDB.

Variables dans Hive

Vous pouvez inclure des variables dans vos scripts en utilisant le symbole dollar et des accolades.

```
add jar ${LIB}/jsonserde.jar
```

Vous transmettez les valeurs de ces variables à Hive sur la ligne de commande à l'aide du paramètre `-d`, comme dans l'exemple suivant :

```
-d LIB=s3://elasticmapreduce/samples/hive-ads/lib
```

Vous pouvez également transmettre les valeurs via des étapes qui exécutent des scripts Hive.

Pour transmettre des valeurs de variables dans des étapes Hive à l'aide de la console

1. Ouvrez la EMR console Amazon à l'[adresse https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr).
2. Choisissez Créer un cluster.
3. Dans la section Étapes, pour Add Step (Ajouter étapes), choisissez Programme Hive dans la liste et Configurer et ajouter.
4. Dans la boîte de dialogue Ajouter une étape, spécifiez les paramètres en utilisant le tableau suivant comme guide, puis choisissez Ajouter.

Champ	Action
Emplacement S3 du script*	Spécifiez l'URL d'emplacement de votre script dans Amazon S3. La valeur doit être sous la forme <i>BucketName /path/ScriptName</i> . Par exemple : <code>s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q</code> .
Emplacement S3 d'entrée	Spécifiez éventuellement l'URL d'emplacement de vos fichiers d'entrée dans Amazon S3. La valeur doit être sous la forme <i>BucketName /path/</i> . S'il est spécifié, il sera transmis au script Hive sous forme de paramètre nommé INPUT. Par exemple : <code>s3://elasticmapreduce/samples/hive-ads/tables/</code> .
Emplacement S3 de sortie	Spécifiez éventuellement l'URL d'endroit où vous souhaitez que la sortie soit stockée dans Amazon S3. La valeur doit être sous la forme <i>BucketName /path</i> . S'il est spécifié, il sera transmis au script Hive sous forme de paramètre nommé OUTPUT. Par exemple : <code>s3://mybucket/hive-ads/output/</code> .
Arguments	<p>Le cas échéant, entrez une liste d'arguments (chaînes séparées par des espaces) à transmettre à Hive. Si vous avez défini une variable de chemin dans votre script Hive nommée <code> \${SAMPLE}</code>, par exemple :</p> <pre>CREATE EXTERNAL TABLE logs (requestBeginTime STRING, requestEndTime STRING, hostname STRING) PARTITIONED BY (dt STRING) \ ROW FORMAT serde 'com.amazon.elasticmapreduce.JsonSerde' WITH SERDEPROPERTIES ('paths'='requestBeginTime, requestEndTime, hostname') LOCATION '\${SAMPLE}/tables/impressions';</pre> <p>Pour transmettre une valeur pour la variable, saisissez ce qui suit dans la fenêtre Arguments :</p> <pre>-d SAMPLE=s3://elasticmapreduce/samples/hive-ads/</pre>

Champ	Action
Action sur échec	<p>Cela détermine ce que le cluster fait en réponse à toute erreur. Les valeurs possibles pour ce paramètre sont :</p> <ul style="list-style-type: none"> • Arrêter le cluster : si l'étape échoue, arrêtez le cluster. Si la protection de terminaison est activée sur le cluster AND Keep Alive, elle ne s'arrêtera pas. • Annuler et attendre : si l'étape échoue, annulez les étapes restantes. Si le cluster a la persistance activée, le cluster ne s'arrêtera pas. • Continuer : si l'étape échoue, passez à l'étape suivante.

5. Sélectionnez les valeurs si nécessaire, puis choisissez Créer un cluster.

Pour transmettre des valeurs variables aux étapes Hive à l'aide du AWS CLI

Pour transmettre des valeurs variables aux étapes Hive à l'aide du paramètre AWS CLI, utilisez le `--steps` paramètre et incluez une liste d'arguments.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

```
aws emr create-cluster --name "Test cluster" --release-label emr-7.2.0 \
--applications Name=Hive Name=Pig --use-default-roles --ec2-attributes
KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=Hive,Name="Hive Program",ActionOnFailure=CONTINUE,Args=[-f,s3://
elasticmapreduce/samples/hive-ads/libs/response-time-stats.q,-d,INPUT=s3://
elasticmapreduce/samples/hive-ads/tables,-d,OUTPUT=s3://mybucket/hive-ads/output/,
-d,SAMPLE=s3://elasticmapreduce/samples/hive-ads/]
```

Pour plus d'informations sur l'utilisation EMR des commandes Amazon dans le AWS CLI, consultez <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Pour transmettre des valeurs variables aux étapes Hive à l'aide de Java SDK

- L'exemple suivant montre comment transmettre des variables à des étapes à l'aide du SDK. Pour plus d'informations, voir [Classe StepFactory](#) dans la AWS SDK for Java API référence.

```
StepFactory stepFactory = new StepFactory();

StepConfig runHive = new StepConfig()
    .withName("Run Hive Script")
    .withActionOnFailure("TERMINATE_JOB_FLOW")
    .withHadoopJarStep(stepFactory.newRunHiveScriptStep("s3://mybucket/script.q",
        Lists.newArrayList("-d", "LIB= s3://elasticmapreduce/samples/hive-ads/lib")));
```

Requêtes Amazon EMR Hive pour prendre en charge des schémas DynamoDB partiels

Amazon EMR Hive offre une flexibilité maximale lors de l'interrogation des tables DynamoDB en vous permettant de spécifier un sous-ensemble de colonnes sur lequel vous pouvez filtrer les données, plutôt que de demander à votre requête d'inclure toutes les colonnes. Cette technique de requête de schéma partiel est efficace lorsque vous avez un schéma de base de données fragmenté et que vous souhaitez filtrer des enregistrements en fonction de quelques colonnes, par exemple le filtrage des horodatages.

L'exemple suivant montre comment utiliser une requête Hive pour :

- Créez une table DynamoDB.
- Sélectionner un sous-ensemble d'éléments (lignes) dans DynamoDB et réduire les données à certaines colonnes.
- Copier les données qui en résultent dans Amazon S3.

```
DROP TABLE dynamodb;
DROP TABLE s3;

CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, fullColumn
map<String, String>)
    STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
    TBLPROPERTIES (
        "dynamodb.table.name" = "myTable",
```



```

"dynamodb.throughput.read.percent" = ".1000",
"dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");

CREATE EXTERNAL TABLE s3(map<String, String>)
  ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
  LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3 SELECT item fullColumn FROM dynamodb WHERE recordTimeStamp <
"2012-01-01";

```

La table suivante illustre la syntaxe de requête pour la sélection de n'importe quelle combinaison d'éléments de DynamoDB.

Exemple de requête	Description du résultat
<code>SELECT * FROM <i>table_name</i> ;</code>	Sélectionne tous les éléments (lignes) dans une table donnée et inclut des données à partir de toutes les colonnes disponibles pour ces éléments.
<code>SELECT * FROM <i>table_name</i> WHERE <i>field_name</i> =<i>value</i>;</code>	Sélectionne certains éléments (lignes) dans une table donnée et inclut des données à partir de toutes les colonnes disponibles pour ces éléments.
<code>SELECT <i>column1_name</i> , <i>column2_name</i> , <i>column3_name</i> FROM <i>table_name</i> ;</code>	Sélectionne tous les éléments (lignes) dans une table donnée et inclut des données à partir de certaines colonnes disponibles pour ces éléments.
<code>SELECT <i>column1_name</i> , <i>column2_name</i> , <i>column3_name</i> FROM <i>table_name</i> WHERE <i>field_name</i> =<i>value</i>;</code>	Sélectionne certains éléments (lignes) dans une table donnée et inclut des données à partir de certaines colonnes disponibles pour ces éléments.

Copie des données entre les tables DynamoDB dans différentes régions AWS

Amazon EMR Hive fournit une `dynamodb.region` propriété que vous pouvez définir par table DynamoDB. Lorsque `dynamodb.region` est définie différemment sur deux tables, toute données que vous copiez entre les tables se produit automatiquement entre les régions spécifiées.

L'exemple suivant vous montre comment créer une table DynamoDB avec un script Hive qui définit la propriété `dynamodb.region` :

Note

Les propriétés de région par table remplacent les propriétés Hive globales.

```
CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, map<String,
String> fullColumn)
  STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
  TBLPROPERTIES (
    "dynamodb.table.name" = "myTable",
    "dynamodb.region" = "eu-west-1",
    "dynamodb.throughput.read.percent" = ".1000",
    "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

Définition des valeurs de débit de DynamoDB par table

Amazon EMR Hive vous permet de définir le `readThroughputPercent` DynamoDB `writeThroughputPercent` et les paramètres par table dans la définition de la table. Le script Amazon EMR Hive suivant montre comment définir les valeurs de débit. Pour plus d'informations sur les valeurs de débit DynamoDB, consultez [Spécification d'exigences de lecture et d'écriture pour des tables](#).

```
CREATE EXTERNAL TABLE dynamodb(hashKey STRING, recordTimeStamp BIGINT, map<String,
String> fullColumn)
  STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
  TBLPROPERTIES (
    "dynamodb.table.name" = "myTable",
    "dynamodb.throughput.read.percent" = ".4",
    "dynamodb.throughput.write.percent" = "1.0",
    "dynamodb.column.mapping" = "hashKey:HashKey,recordTimeStamp:RangeKey");
```

Configuration d'un metastore externe pour Hive

Par défaut, Hive enregistre les informations du métastore dans une SQL base de données My sur le système de fichiers du nœud principal. Le metastore contient une description de la table et des données sous-jacentes sur lesquelles il est construit, notamment les noms de partition, les types de données, etc. Lorsqu'un cluster est suspendu, tous les nœuds de cluster sont arrêtés, y compris le nœud primaire. Lorsque cela arrive, les données locales sont perdues, car les systèmes de fichiers

du nœud utilisent un magasin éphémère. Pour que le metastore persiste, vous devez créer un metastore externe qui existe en dehors du cluster.

Vous avez deux options pour un metastore externe :

- AWS Glue Data Catalog (Amazon EMR version 5.8.0 ou ultérieure uniquement).

Pour de plus amples informations, veuillez consulter [Utiliser le catalogue de données AWS Glue comme métastore pour Hive](#).

- Amazon RDS ou Amazon Aurora.

Pour de plus amples informations, veuillez consulter [Utilisation d'une base de SQL données My Database externe ou d'Amazon Aurora](#).

Note

Si vous utilisez Hive 3 et que vous rencontrez trop de connexions au métastore Hive, configurez le paramètre `datanucleus.connectionPool.maxPoolSize` pour qu'il ait une valeur inférieure ou augmentez le nombre de connexions que le serveur de base de données peut gérer. L'augmentation du nombre de connexions est due à la façon dont Hive calcule le nombre maximum de JDBC connexions. Pour calculer la valeur optimale en termes de performances, consultez la section [Propriétés de configuration de Hive](#).

Utiliser le catalogue de données AWS Glue comme métastore pour Hive

À l'aide de la EMR version 5.8.0 ou ultérieure d'Amazon, vous pouvez configurer Hive pour utiliser le catalogue de données AWS Glue comme métastore. Cette configuration est recommandée si vous avez besoin d'un metastore persistant ou d'un metastore partagé par différents clusters, services, applications ou comptes AWS .

AWS Glue est un service d'extraction, de transformation et de chargement (ETL) entièrement géré qui permet de classer vos données, de les nettoyer, de les enrichir et de les déplacer de manière fiable entre différents magasins de données de manière simple et rentable. Le catalogue de données AWS Glue fournit un référentiel de métadonnées unifié couvrant une variété de sources de données et de formats de données, s'intégrant à Amazon EMR ainsi qu'à AmazonRDS, Amazon Redshift, Redshift Spectrum, Athena et à toute application compatible avec le métastore Apache Hive. AWS Les robots d'exploration Glue peuvent automatiquement déduire un schéma à partir des données

sources dans Amazon S3 et stocker les métadonnées associées dans le catalogue de données. Pour plus d'informations sur le catalogue de données, consultez la section [Remplissage du catalogue de données AWS Glue](#) dans le guide du développeur de AWS Glue.

Des frais distincts s'appliquent pour AWS Glue. Il existe un tarif mensuel pour le stockage et l'accès aux métadonnées du catalogue de données, un taux horaire facturé par minute pour les ETL tâches AWS Glue et le temps d'exécution du robot d'exploration, et un taux horaire facturé par minute pour chaque point de terminaison de développement provisionné. Le catalogue de données vous permet de stocker jusqu'à un million d'objets sans aucuns frais. Si vous stockez plus d'un million d'objets, vous êtes facturé USD 1\$ pour 100 000 objets de plus d'un million. Dans le catalogue de données, sont considérés comme des objets une table, une partition ou une base de données. Pour de plus amples informations, consultez [Tarification Glue](#).

Important

Si vous avez créé des tables à l'aide d'Amazon Athena ou d'Amazon Redshift Spectrum avant le 14 août 2017, les bases de données et les tables sont stockées dans un catalogue géré par Athena, distinct du catalogue de données Glue. AWS Pour intégrer Amazon EMR à ces tables, vous devez passer au catalogue de données AWS Glue. Pour plus d'informations, consultez la section [Mise à niveau vers le catalogue de données AWS Glue](#) dans le guide de l'utilisateur Amazon Athena.

Spécifier le catalogue de données AWS Glue comme métastore

Vous pouvez spécifier le catalogue de données AWS Glue comme métastore à l'aide de AWS Management Console AWS CLI, ou Amazon. EMR API Lorsque vous utilisez le CLI ou API, vous utilisez la classification de configuration de Hive pour spécifier le catalogue de données. En outre, avec Amazon EMR 5.16.0 et versions ultérieures, vous pouvez utiliser la classification de configuration pour spécifier un catalogue de données dans un autre. Compte AWS Lorsque vous utilisez la console, vous pouvez spécifier le catalogue de données à l'aide des Options avancées ou des Options rapides.

Console

Pour spécifier AWS Glue Data Catalog comme métastore Hive avec la console

1. Connectez-vous à la AWS Management Console EMR console Amazon et ouvrez-la à l'adresse <https://console.aws.amazon.com/emr>.

2. Sous EMREC2Activé dans le volet de navigation de gauche, choisissez Clusters, puis Create cluster.
3. Sous Ensemble d'applications, choisissez Core Hadoop ou HBaseCustom. Si vous personnalisez votre cluster, assurez-vous de sélectionner Hive ou HCatalog comme l'une de vos applications.
4. Sous Paramètres du catalogue de données AWS Glue, cochez la case Utiliser les métadonnées des tables Hive.
5. Choisissez toutes les autres options qui s'appliquent à votre cluster.
6. Pour lancer cluster, choisissez Créer un cluster.

CLI

Pour spécifier le catalogue de données AWS Glue en tant que métastore Hive avec AWS CLI

Pour plus d'informations sur la spécification d'une classification de configuration à l'aide du AWS CLI et EMRAPI, consultez [Configuration des applications](#).

- Spécifiez la valeur de `hive.metastore.client.factory.class` à l'aide de la classification de configuration `hive-site` comme illustré dans l'exemple suivant :

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
"com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

Dans les EMR versions 5.28.0, 5.28.1, 5.29.0 ou 6.x, si vous créez un cluster en utilisant le catalogue de données AWS Glue comme métastore, définissez le sur. `hive.metastore.schema.validation false` Cela empêche Hive et HCatalog de valider le schéma du métastore par rapport à My. SQL Sans cette configuration, le groupe d'instances principal sera suspendu après la reconfiguration sur Hive ou. HCatalog

```
[
```

```
{
  "Classification": "hive-site",
  "Properties": {
    "hive.metastore.client.factory.class":
    "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
    "hive.metastore.schema.validation": "false"
  }
}
```

Si vous disposez déjà d'un cluster dans les EMR versions 5.28.0, 5.28.1 ou 5.29.0, vous pouvez définir le groupe d'instances principal à l'aide des informations suivantes :

```
hive.metastore.schema.validation false
```

```
Classification = hive-site
Property       = hive.metastore.schema.validation
Value         = false
```

Pour spécifier un catalogue de données dans un autre AWS compte, ajoutez la `hive.metastore.glue.catalogid` propriété comme indiqué dans l'exemple suivant. Remplacez *acct-id* par le compte AWS du catalogue de données.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.schema.validation": "false",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

IAM autorisations

Le profil d'EC2 instance d'un cluster doit disposer d'IAM autorisations pour les actions AWS Glue. En outre, si vous activez le chiffrement pour les objets AWS Glue Data Catalog, le rôle doit également être autorisé à chiffrer, déchiffrer et générer les objets AWS KMS key utilisés pour le chiffrement.

Autorisations pour les actions AWS Glue

Si vous utilisez le profil d'EC2 instance par défaut pour Amazon EMR, aucune action n'est requise. La politique `AmazonElasticMapReduceforEC2Role` gérée attachée au `EMR_EC2_DefaultRole` autorise toutes les actions AWS Glue nécessaires. Toutefois, si vous spécifiez un profil d'EC2 instance personnalisé et des autorisations, vous devez configurer les actions AWS Glue appropriées. Utilisez la stratégie gérée `AmazonElasticMapReduceforEC2Role` comme point de départ. Pour plus d'informations, consultez la section [Rôle de service pour les EC2 instances de cluster \(profil d'EC2 instance\)](#) dans le guide EMR de gestion Amazon.

Autorisations pour chiffrer et déchiffrer le catalogue de données Glue AWS

Votre profil d'instance a besoin d'autorisations pour chiffrer et déchiffrer les données au moyen de votre clé. Vous n'avez pas besoin de configurer ces autorisations si les deux déclarations suivantes s'appliquent :

- Vous activez le chiffrement des objets du catalogue de données AWS Glue à l'aide de clés gérées pour AWS Glue.
- Vous utilisez un cluster Compte AWS identique au AWS Glue Data Catalog.

Dans le cas contraire, vous devez ajouter l'instruction suivante à la politique d'autorisation attachée à votre profil d'EC2 instance.

```
[
  {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Action": [
          "kms:Decrypt",
          "kms:Encrypt",
          "kms:GenerateDataKey"
        ]
      }
    ]
  }
]
```

```

        "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
    }
]
}
]

```

Pour plus d'informations sur le chiffrement du catalogue de données AWS Glue, consultez la section [Chiffrer votre catalogue de données](#) dans le guide du développeur de AWS Glue.

Autorisations basées sur les ressources

Si vous utilisez AWS Glue conjointement avec Hive, Spark ou Presto sur Amazon, EMR AWS Glue prend en charge les politiques basées sur les ressources afin de contrôler l'accès aux ressources du catalogue de données. Ces ressources comprennent les bases de données, les tables, les connexions et les fonctions définies par l'utilisateur. Pour plus d'informations, consultez [Politiques de ressources AWS Glue](#) dans le Guide du développeur AWS Glue.

Lorsque vous utilisez des politiques basées sur les ressources pour limiter l'accès à AWS Glue depuis AmazonEMR, le principal que vous spécifiez dans la politique d'autorisation doit être le rôle ARN associé au profil d'EC2instance spécifié lors de la création d'un cluster. Par exemple, pour une politique basée sur les ressources attachée à un catalogue, vous pouvez spécifier le rôle du rôle de service par défaut ARN pour les instances de clusterEC2, *EMR_EC2_DefaultRole* comme lePrincipal, en utilisant le format illustré dans l'exemple suivant :

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

Le *acct-id* peut être différent de l'identifiant du compte AWS Glue. Cela permet d'accéder à partir de EMR clusters appartenant à différents comptes. Vous pouvez spécifier plusieurs principaux, chacun provenant d'un compte différent.

Considérations relatives à l'utilisation du catalogue de données AWS Glue

Tenez compte des éléments suivants lorsque vous utilisez le catalogue de données AWS Glue comme métastore avec Hive :

- L'ajout d'un auxiliaire JARs à l'aide du shell Hive n'est pas pris en charge. Pour contourner le problème, utilisez la classification `hive-site` de configuration pour définir la `hive.aux.jars.path` propriété, qui ajoute un auxiliaire JARs au chemin de classe Hive.
- Les [transactions Hive](#) ne sont pas prises en charge.

- Il n'est pas possible de renommer des tables depuis AWS Glue.
- Lorsque vous créez une table Hive sans spécifier de LOCATION, les données de la table sont stockées à l'emplacement spécifié par la propriété `hive.metastore.warehouse.dir`. Par défaut, il s'agit d'un emplacement dans HDFS. Si un autre cluster doit accéder à la table, il échoue, sauf s'il dispose d'autorisations adéquates sur le cluster qui a créé la table. En outre, le HDFS stockage étant transitoire, si le cluster se termine, les données de la table sont perdues et la table doit être recréée. Nous vous recommandons de spécifier un LOCATION dans Amazon S3 lorsque vous créez une table Hive à l'aide de AWS Glue. Vous pouvez également utiliser la classification de configuration `hive-site` pour spécifier un emplacement dans Amazon S3 pour `hive.metastore.warehouse.dir`, qui s'applique à toutes les tables Hive. Si une table est créée dans un HDFS emplacement et que le cluster qui l'a créée est toujours actif, vous pouvez mettre à jour l'emplacement de la table vers Amazon S3 depuis AWS Glue. Pour plus d'informations, consultez [Working with Tables on the AWS Glue Console](#) dans le AWS Glue Developer Guide.
- Les valeurs de partition contenant des guillemets et des apostrophes ne sont pas prises en charge, par exemple `PARTITION (owner="Doe 's")`.
- Les [statistiques sur les colonnes](#) sont prises en charge à partir de la version `emr-5.31.0`.
- L'utilisation de l'[autorisation Hive](#) n'est pas prise en charge. Comme alternative, envisagez d'utiliser les [politiques basées sur les ressources AWS Glue](#). Pour plus d'informations, consultez [Utiliser des politiques basées sur les ressources pour le catalogue de données Amazon EMR Access to AWS Glue](#).
- Les [contraintes Hive](#) ne sont pas prises en charge.
- L'[optimisation basée sur les coûts dans Hive](#) n'est pas prise en charge.
- Le paramètre `hive.metastore.partition.inherit.table.properties` n'est pas pris en charge.
- L'utilisation des constantes de métastore suivantes n'est pas prise en charge :
BUCKET_COUNT, BUCKET_FIELD_NAME, DDL_TIME, FIELD_TO_DIMENSION,
FILE_INPUT_FORMAT, FILE_OUTPUT_FORMAT, HIVE_FILTER_FIELD_LAST_ACCESS,
HIVE_FILTER_FIELD_OWNER, HIVE_FILTER_FIELD_PARAMS, IS_ARCHIVED,
META_TABLE_COLUMNS, META_TABLE_COLUMN_TYPES, META_TABLE_DB,
META_TABLE_LOCATION, META_TABLE_NAME, META_TABLE_PARTITION_COLUMNS,
META_TABLE_SERDE, META_TABLE_STORAGE, ORIGINAL_LOCATION.
- Lorsque vous utilisez une expression de prédicat, les valeurs explicites doivent se trouver sur le côté droit de l'opérateur de comparaison, sinon les requêtes risquent d'échouer.

- **Correct** : `SELECT * FROM mytable WHERE time > 11`
- **Incorrect** : `SELECT * FROM mytable WHERE 11 > time`
- EMR Les versions 5.32.0 et 6.3.0 et ultérieures d'Amazon prennent en charge l'utilisation de fonctions définies par l'utilisateur (UDFs) dans les expressions de prédicat. Lorsque vous utilisez des versions antérieures, vos requêtes peuvent échouer en raison de la manière dont Hive essaie d'optimiser l'exécution des requêtes.
- Les [tables temporaires](#) ne sont pas prises en charge.
- Nous vous recommandons de créer des tableaux à l'aide d'applications via Amazon EMR plutôt que de les créer directement avec AWS Glue. La création d'une table via AWS Glue peut entraîner l'absence de champs obligatoires et des exceptions de requête.
- Dans la version EMR 5.20.0 ou ultérieure, l'élagage parallèle des partitions est activé automatiquement pour Spark et Hive lorsque AWS Glue Data Catalog est utilisé comme métastore. Cette modification réduit considérablement le temps de planification des requêtes en exécutant plusieurs requêtes en parallèle pour récupérer des partitions. Le nombre total de segments pouvant être exécutés simultanément est compris entre 1 et 10. La valeur par défaut est 5, ce qui est recommandé. Vous pouvez le modifier en spécifiant la propriété `aws.glue.partition.num.segments` dans la classification de configuration `hive-site`. En cas de limitation, vous pouvez désactiver la fonctionnalité en remplaçant la valeur par 1. Pour en savoir plus, consultez [Structure d'un segment AWS Glue](#).

Utilisation d'une base de SQL données My Database externe ou d'Amazon Aurora

Pour utiliser une base de SQL données My database externe ou Amazon Aurora comme métastore Hive, vous devez remplacer les valeurs de configuration par défaut du métastore dans Hive afin de spécifier l'emplacement de la base de données externe, que ce soit sur une instance Amazon My RDS ou SQL Amazon Aurora P. `ostgreSQLInstance`

Note

Hive ne prend pas en charge et n'empêche pas l'accès simultané en écriture aux tables du métastore. Si vous partagez des informations du métastore entre deux clusters, vous devez veiller à ne pas écrire dans la même table du métastore simultanément, sauf si vous écrivez dans des partitions différentes de la même table du métastore.

La procédure suivante vous montre comment remplacer les valeurs de configuration par défaut pour l'emplacement du metastore Hive et démarrer un cluster en utilisant l'emplacement du metastore reconfiguré.

Pour créer un métastore situé en dehors du cluster EMR

1. Créez une base de SQL données My SQL ou Aurora Postgre. Si vous utilisez PostgreSQL, vous devez le configurer après avoir provisionné votre cluster. Seul My SQL est pris en charge lors de la création du cluster. Pour plus d'informations sur les différences entre Aurora My SQL et Aurora PostgreSQL, consultez [Présentation d'Amazon Aurora My SQL](#) et [Utilisation d'Amazon Aurora SQL Postgre](#). Pour plus d'informations sur la création d'une RDS base de données Amazon en général, consultez <https://aws.amazon.com/rds/>.
2. Modifiez vos groupes de sécurité pour autoriser JDBC les connexions entre votre base de données et le groupe de sécurité ElasticMapReduce-Master. Pour plus d'informations sur la façon de modifier vos groupes de sécurité pour l'accès, consultez la section [Travailler avec les groupes de sécurité EMR gérés par Amazon](#).
3. Définissez les valeurs JDBC de configuration dans `hive-site.xml` :

⚠ Important

Si vous fournissez des informations sensibles, telles que des mots de passe, à la EMR configuration AmazonAPI, ces informations sont affichées pour les comptes disposant d'autorisations suffisantes. Si vous craignez que ces informations ne soient affichées à d'autres utilisateurs, créez le cluster avec un compte administratif et limitez l'accès des autres IAM utilisateurs (utilisateurs ou détenteurs d'informations d'identification déléguées) aux services du cluster en créant un rôle qui refuse explicitement les autorisations relatives à la `elasticmapreduce:DescribeCluster` API clé.

- a. Créez un fichier de configuration appelé `hiveConfiguration.json` contenant les modifications apportées à `hive-site.xml` comme illustré dans l'exemple suivant.

Remplacez *hostname* avec l'adresse DNS de votre RDS instance Amazon exécutant la base de données, et *username* and *password* avec les informations d'identification de votre base de données. Pour plus d'informations sur la connexion aux instances de base de données My SQL et Aurora, consultez les sections [Connexion à une instance de base de données exécutant le moteur de SQL base de données My](#) et [Connexion](#)

à un cluster de base de données Athena dans le guide de RDS l'utilisateur Amazon.

`javax.jdo.option.ConnectionURL` est la chaîne de JDBC connexion d'un JDBC métastore. `javax.jdo.option.ConnectionDriverName` est le nom de classe de pilote d'un JDBC métastore.

Mes SQL JDBC pilotes sont installés par AmazonEMR.

La propriété value ne peut pas contenir d'espaces ni de retours chariot. Elle doit figurer entièrement sur une seule ligne.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "javax.jdo.option.ConnectionURL": "jdbc:mysql://hostname:3306/hive?
createDatabaseIfNotExist=true",
      "javax.jdo.option.ConnectionDriverName": "org.mariadb.jdbc.Driver",
      "javax.jdo.option.ConnectionUserName": "username",
      "javax.jdo.option.ConnectionPassword": "password"
    }
  }
]
```

- b. Référez le `hiveConfiguration.json` fichier lorsque vous créez le cluster, comme indiqué dans la AWS CLI commande suivante. Le fichier est stocké localement dans cette commande. Vous pouvez également charger le fichier sur Amazon S3 et le référencer depuis cet endroit, par exemple, `s3://DOC-EXAMPLE-BUCKET/hiveConfiguration.json`.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --release-label emr-7.2.0 --instance-type m5.xlarge --
instance-count 2 \
--applications Name=Hive --configurations file://hiveConfiguration.json --use-
default-roles
```

4. Connectez-vous au nœud primaire de votre cluster.

Pour plus d'informations sur la connexion au nœud principal, consultez la section [Se connecter au nœud principal SSH à l'aide](#) du guide EMR de gestion Amazon.

5. Créez vos tables Hive en spécifiant l'emplacement sur Amazon S3 en entrant une commande similaire à :

```
CREATE EXTERNAL TABLE IF NOT EXISTS table_name
(
  key int,
  value int
)
LOCATION s3://DOC-EXAMPLE-BUCKET/hdfs/
```

6. Ajoutez votre script Hive au cluster en cours d'exécution.

Votre cluster Hive s'exécute à l'aide du metastore situé sur Amazon RDS. Lancez tous les clusters Hive supplémentaires qui partagent ce metastore en spécifiant l'emplacement du metastore.

Utilisation du pilote Hive JDBC

Vous pouvez utiliser des outils populaires d'aide à la décision, comme Microsoft Excel, MicroStrategy, QlikView et Tableau, avec Amazon EMR pour explorer et visualiser vos données. Un grand nombre de ces outils ont besoin d'un pilote JDBC (Java DataBase Connectivity) ou ODBC (Open DataBase Connectivity). Amazon EMR prend en charge la connectivité JDBC et ODBC.

L'exemple ci-dessous explique comment utiliser SQL Workbench/J en tant que client SQL pour vous connecter à un cluster Hive dans Amazon EMR. Pour obtenir des pilotes supplémentaires, consultez [Utilisation des outils d'aide à la décision avec Amazon EMR](#).

Avant d'installer et d'utiliser SQL Workbench/J, téléchargez le package du pilote et installez ce dernier. Les pilotes inclus dans le package prennent en charge les versions Hive disponibles dans Amazon EMR versions 4.0 et versions ultérieures. Pour obtenir des notes de mise à jour et une documentation détaillées, consultez la documentation PDF incluse dans le package.

- Téléchargement du dernier package de pilotes Hive JDBC

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/>

- Anciennes versions du pilote JDBC Hive

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/>

Pour installer et configurer SQL Workbench

1. Téléchargez le client SQL Workbench/J pour votre système d'exploitation à l'adresse <http://www.sql-workbench.net/downloads.html>.
2. Installez SQL Workbench/J. Pour plus d'informations, consultez [Installing and starting SQL Workbench/J](#) dans le manuel de l'utilisateur SQL Workbench/J.
3. Utilisateurs Linux, Unix et Mac OS X : dans une session de terminal, créez un tunnel SSH vers le nœud principal de votre cluster à l'aide de la commande suivante. Remplacez *master-public-dns-name* par le nom DNS public du nœud principal et *path-to-key-file* par l'emplacement et le nom de votre fichier de clé privée Amazon EC2 (.pem).

```
ssh -o ServerAliveInterval=10 -i path-to-key-file -N -L 10000:localhost:10000  
hadoop@master-public-dns-name
```

Utilisateurs Windows : Dans une session PuTTY, créez un tunnel SSH pour le nœud principal de votre cluster (à l'aide du réacheminement de port local) avec 10000 pour Port source et *master-public-dns-name*:10000 pour Destination. Remplacez *master-public-dns-name* par le nom DNS public du nœud maître.

4. Ajoutez le pilote JDBC à SQL Workbench.
 - a. Dans la boîte de dialogue Select Connection Profile (Sélectionner le profil de connexion), cliquez sur Manage Drivers (Gérer les pilotes).
 - b. Cliquez sur l'icône Create a new entry (Créer une nouvelle entrée) (page blanche).
 - c. Dans le champ Name (Nom), saisissez **Hive JDBC**.
 - d. Pour Library (Bibliothèque), cliquez sur l'icône Select the JAR file(s) (Sélectionner le(s) fichier(s) JAR).
 - e. Accédez à l'emplacement contenant les pilotes extraits. Sélectionnez les pilotes inclus dans la version du package de pilotes JDBC que vous avez téléchargée, puis cliquez sur Ouvrir.

Par exemple, votre package de pilote JDBC peut inclure les fichiers JAR suivants.

```
hive_metastore.jar  
hive_service.jar  
HiveJDBC41.jar
```

```
libfb303-0.9.0.jar  
libthrift-0.9.0.jar  
log4j-1.2.14.jar  
ql.jar  
slf4j-api-1.5.11.jar  
slf4j-log4j12-1.5.11.jar  
TCLIServiceClient.jar  
zookeeper-3.4.6.jar
```

- f. Dans la boîte de dialogue Please select one driver (Veuillez sélectionner un pilote), sélectionnez `com.amazon.hive.jdbc41.HS2Driver`, OK.
5. Lorsque vous revenez à la boîte de dialogue Manage Drivers (Gérer les pilotes) vérifiez que le champ Classname (Nom de classe) est renseigné et cliquez sur OK.
6. Lorsque vous revenez à la boîte de dialogue Select Connection Profile (Sélectionner le profil de connexion), vérifiez que le champ Driver (Pilote) est défini sur Hive JDBC et fournissez la chaîne de connexion JDBC suivante dans le champ URL : `jdbc:hive2://localhost:10000/default`.
7. Sélectionnez OK pour vous connecter. Une fois la connexion établie, les détails de connexion apparaissent en haut de la fenêtre SQL Workbench/J.

Pour plus d'informations sur l'utilisation de Hive et de l'interface JDBC, consultez [HiveClient](#) et [HiveJDBCInterface](#) dans la documentation Apache Hive.

Amélioration des performances de Hive

Amazon EMR propose des fonctionnalités permettant d'optimiser les performances lors de l'utilisation de Hive pour interroger, lire et écrire des données enregistrées dans Amazon S3.

S3 Select peut améliorer les performances des requêtes CSV et JSON des fichiers dans certaines applications en « transférant » le traitement vers Amazon S3.

Le committer optimisé EMRFS S3 est une alternative à la [OutputCommitter](#) classe, qui élimine les opérations de liste et de renommage afin d'améliorer les performances lors de l'écriture de fichiers utilisés par Amazon S3. EMRFS

Rubriques

- [Activation du validateur optimisé Hive EMRFS S3](#)
- [Utilisation de S3 Select avec Hive pour améliorer les performances](#)

- [MSCKOptimisation](#)

Activation du validateur optimisé Hive EMRFS S3

Le Hive EMRFS S3 Optimized Committer est une méthode alternative permettant à EMR Hive d'écrire des fichiers pour insérer des requêtes lors de l'utilisation. EMRFS Le Committer élimine les opérations de liste et de renommage effectuées sur Amazon S3 et améliore les performances de l'application. La fonctionnalité est disponible à partir des versions EMR 5.34 et EMR 6.5.

Activation du contributeur

Si vous souhaitez activer EMR Hive pour valider les données par défaut pour toutes les tables gérées et externes par Hive, utilisez la `hive-site` configuration suivante dans les clusters EMR 6.5.0 ou EMR 5.34.0. `HiveEMRFSOptimizedCommitter`

```
[
  {
    "classification": "hive-site",
    "properties": {
      "hive.blobstore.use.output-committer": "true"
    }
  }
]
```

Note

N'activez pas cette fonction lorsque `hive.exec.parallel` est défini sur `true`.

Limites

Les restrictions de base suivantes s'appliquent aux balises :

- L'activation de Hive pour fusionner automatiquement de petits fichiers n'est pas prise en charge. La logique de validation Hive par défaut sera utilisée même lorsque le validateur optimisé est activé.
- Les ACID tables Hive ne sont pas prises en charge. La logique de validation Hive par défaut sera utilisée même lorsque le validateur optimisé est activé.

- La nomenclature des noms de fichiers pour les fichiers écrits est passée de Hive `<task_id>_<attempt_id>_<copy_n>` à `<task_id>_<attempt_id>_<copy_n>_<query_id>`. Par exemple, un fichier nommé `s3://warehouse/table/partition=1/000000_0` sera remplacé par `s3://warehouse/table/partition=1/000000_0-hadoop_20210714130459_ba7c23ec-5695-4947-9d98-8a40ef759222-1.query_id`. Voici une combinaison du nom d'utilisateur, de l'horodatage et d'UUID.
- Lorsque des partitions personnalisées se trouvent sur des systèmes de fichiers différents (HDFS/S3), cette fonctionnalité est automatiquement désactivée. La logique de validation Hive par défaut sera utilisée lorsqu'elle sera activée.

Utilisation de S3 Select avec Hive pour améliorer les performances

Avec les EMR versions 5.18.0 et ultérieures d'Amazon, vous pouvez utiliser [S3 Select](#) avec Hive sur Amazon. EMR S3 Select permet aux applications de ne récupérer qu'un sous-ensemble de données d'un objet. Pour AmazonEMR, le travail de calcul consistant à filtrer de grands ensembles de données à des fins de traitement est « transféré » du cluster vers Amazon S3, ce qui peut améliorer les performances de certaines applications et réduire la quantité de données transférées entre Amazon EMR et Amazon S3.

S3 Select est compatible avec les tables Hive basées sur CSV des JSON fichiers et en définissant la variable `s3select.filter` de configuration sur `true` pendant votre session Hive. Pour plus d'informations et d'exemples, consultez [Spécification de S3 Select dans votre code](#).

S3 Select est-il adapté à mon application ?

Nous vous recommandons de tester vos applications avec et sans S3 Select pour voir si son utilisation peut être adaptée à votre application.

Utilisez les consignes suivantes pour déterminer si votre application peut utiliser S3 Select :

- Votre requête exclut plus de la moitié de l'ensemble de données d'origine.
- Vos prédicats de filtre de requête utilisent les colonnes qui ont un type de données pris en charge par Amazon S3 Select. Pour plus d'informations, consultez [Types de données](#) dans le Guide de l'utilisateur Amazon Simple Storage Service.
- Votre connexion réseau entre Amazon S3 et le EMR cluster Amazon offre une bonne vitesse de transfert et une bonne bande passante disponible. Amazon S3 ne compresse pas HTTP les

réponses. La taille des réponses est donc susceptible d'augmenter pour les fichiers d'entrée compressés.

Considérations et restrictions

- Le chiffrement côté serveur Amazon S3 avec les clés de chiffrement fournies par le client (SSE-C) et le chiffrement côté client ne sont pas pris en charge.
- La propriété `AllowQuotedRecordDelimiters` n'est pas prise en charge. Si cette propriété est spécifiée, la requête échoue.
- Seuls CSV les JSON fichiers au format UTF -8 sont pris en charge. Multilignes CSVs et ne JSON sont pas pris en charge.
- Seuls les fichiers non compressés ou gzip ou bzip2 sont pris en charge.
- Les caractères de commentaire ne sont pas pris en charge dans la dernière ligne.
- Des lignes vides à la fin d'un fichier ne sont pas traitées.
- Hive on Amazon EMR prend en charge les types de données primitifs pris en charge par S3 Select. Pour plus d'informations, consultez [Types de données](#) dans le Guide de l'utilisateur Amazon Simple Storage Service.

Spécification de S3 Select dans votre code

Pour utiliser S3 Select dans votre table Hive, créez la table en spécifiant `com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat` comme nom de classe `INPUTFORMAT` et spécifiez une valeur pour la propriété `s3select.format` à l'aide de la clause `TBLPROPERTIES`.

S3 Select est désactivé par défaut lorsque vous exécutez des requêtes. Activez S3 Select en définissant `s3select.filter` sur `true` dans votre session Hive, comme illustré ci-dessous. Les exemples ci-dessous montrent comment spécifier S3 Select lors de la création d'une table à partir du sous-jacent CSV et de JSON fichiers, puis lors de l'interrogation de la table à l'aide d'une simple instruction `select`.

Exemple `CREATETABLE` déclaration pour une table CSV basée

```
CREATE TABLE mys3selecttable (  
  col1 string,  
  col2 int,
```

```

col3 boolean
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS
INPUTFORMAT
  'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 's3://path/to/mycsvfile/'
TBLPROPERTIES (
  "s3select.format" = "csv",
  "s3select.headerInfo" = "ignore"
);

```

Exemple CREATETABLEdéclaration pour une table JSON basée

```

CREATE TABLE mys3selecttable (
col1 string,
col2 int,
col3 boolean
)
ROW FORMAT SERDE 'org.apache.hive.hcatalog.data.JsonSerDe'
STORED AS
INPUTFORMAT
  'com.amazonaws.emr.s3select.hive.S3SelectableTextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION 's3://path/to/json/'
TBLPROPERTIES (
  "s3select.format" = "json"
);

```

Exemple SELECTTABLEdéclaration

```

SET s3select.filter=true;
SELECT * FROM mys3selecttable WHERE col2 > 10;

```

MSCKOptimisation

Hive stocke une liste de partitions pour chaque table dans son métastore. Toutefois, lorsque des partitions sont directement ajoutées ou supprimées du système de fichiers, le métastore Hive n'est

pas au courant de ces modifications. La [MSCKcommande](#) met à jour les métadonnées de partition dans le métastore Hive pour les partitions directement ajoutées ou supprimées du système de fichiers. La syntaxe de cette commande est la suivante :

```
MSCK [REPAIR] TABLE table_name [ADD/DROP/SYNC PARTITIONS];
```

Hive implémente cette commande comme suit :

1. Hive récupère toutes les partitions de la table dans le métastore. À partir de la liste des chemins de partition qui n'existent pas dans le système de fichiers, crée une liste de partitions à supprimer du métastore.
2. Hive rassemble les chemins de partition présents dans le système de fichiers, les compare à la liste des partitions du métastore et génère une liste des partitions qui doivent être ajoutées au métastore.
3. Hive met à jour le métastore en utilisant le mode ADD, DROP ou SYNC.

Note

Lorsque le métastore contient de nombreuses partitions, l'étape visant à vérifier si une partition n'existe pas dans le système de fichiers prend beaucoup de temps, car l'`exists` API appel du système de fichiers doit être effectué pour chaque partition.

Dans Amazon EMR 6.5.0, Hive a introduit un drapeau appelé.

`hive.emr.optimize.msck.fs.check` Lorsqu'il est activé, cet indicateur permet à Hive de vérifier la présence d'une partition dans la liste des chemins de partition du système de fichiers générée à l'étape 2 ci-dessus au lieu d'effectuer des API appels au système de fichiers. Dans Amazon EMR 6.8.0, Hive a activé cette optimisation par défaut, éliminant ainsi le besoin de définir l'indicateur.

`hive.emr.optimize.msck.fs.check`

Utilisation de Hive Live Long et Process () LLAP

Amazon EMR 6.0.0 prend en charge les fonctionnalités Live Long et Process (LLAP) pour Hive. LLAP utilise des démons persistants dotés d'une mise en cache intelligente en mémoire pour améliorer les performances des requêtes par rapport au mode d'exécution du conteneur Tez par défaut précédent.

Les LLAP démons Hive sont gérés et exécutés en tant que service. YARN Comme un YARN service peut être considéré comme une YARN application de longue durée, certaines des ressources de votre cluster sont dédiées à Hive LLAP et ne peuvent pas être utilisées pour d'autres charges de travail. Pour plus d'informations, consultez [LLAP](#) et [YARNService API](#).

Activer Hive LLAP sur Amazon EMR

Pour activer Hive LLAP sur AmazonEMR, fournissez la configuration suivante lorsque vous lancez un cluster.

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true"
    }
  }
]
```

Pour plus d'informations, consultez [Configuration des applications](#).

Par défaut, Amazon EMR alloue environ 60 % des YARN ressources du cluster aux daemons LLAP Hive. Vous pouvez configurer le pourcentage de YARN ressources du cluster alloué à Hive LLAP et le nombre de tâches et de nœuds principaux à prendre en compte pour l'allocation HiveLLAP.

Par exemple, la configuration suivante démarre Hive LLAP avec trois démons sur trois nœuds de tâches ou nœuds principaux et alloue 40 % des ressources des trois nœuds YARN principaux ou de tâches aux démons Hive. LLAP

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true",
      "hive.llap.percent-allocation": "0.4",
      "hive.llap.num-instances": "3"
    }
  }
]
```

Vous pouvez utiliser les `hive-site` configurations suivantes dans la classification API pour remplacer les paramètres de LLAP ressources par défaut.

Propriété	Description
<code>hive.llap.daemon.yarn.container.mb</code>	Taille totale du conteneur de LLAP démons (en Mo)
<code>hive.llap.daemon.memory.per.instance.mb</code>	Mémoire totale utilisée par les exécuteurs dans le conteneur de LLAP démons (en Mo)
<code>hive.llap.io.memory.size</code>	Taille du cache pour les LLAP entrées/sorties
<code>hive.llap.daemon.num.executors</code>	Nombre d'exécuteurs par démon LLAP

Démarrez Hive LLAP manuellement sur votre cluster

Toutes les dépendances et configurations utilisées par LLAP sont empaquetées dans l'archive LLAP tar dans le cadre du démarrage du cluster. Si cette option LLAP est activée à l'aide de `"hive.llap.enabled": "true"`, nous vous recommandons d'utiliser la EMR reconfiguration d'Amazon pour apporter des modifications de configuration à LLAP.

Sinon, pour toute modification manuelle `hive-site.xml`, vous devez reconstruire l'archive LLAP tar à l'aide de la `hive --service llap` commande, comme le montre l'exemple suivant.

```
# Define how many resources you want to allocate to Hive LLAP

LLAP_INSTANCES=<how many llap daemons to run on cluster>
LLAP_SIZE=<total container size per llap daemon>
LLAP_EXECUTORS=<number of executors per daemon>
LLAP_XMX=<Memory used by executors>
LLAP_CACHE=<Max cache size for IO allocator>

yarn app -enableFastLaunch

hive --service llap \
--instances $LLAP_INSTANCES \
--size ${LLAP_SIZE}m \
--executors $LLAP_EXECUTORS \
```

```
--xmx ${LLAP_XMX}m \  
--cache ${LLAP_CACHE}m \  
--name llap0 \  
--auxhbase=false \  
--startImmediately
```

Vérifier l'état de la ruche LLAP

Utilisez la commande suivante pour vérifier l'état de Hive LLAP via Hive.

```
hive --service llapstatus
```

Utilisez la commande suivante pour vérifier l'état de Hive LLAP en utilisant YARN.

```
yarn app -status (name-of-llap-service)  
  
# example:  
yarn app -status llap0 | jq
```

Démarrer ou arrêter Hive LLAP

Étant donné que Hive LLAP fonctionne en tant que YARN service permanent, vous devez arrêter ou redémarrer le YARN service pour arrêter ou redémarrer Hive LLAP. Les commandes suivantes illustrent ce problème.

```
yarn app -stop llap0  
yarn app -start llap0
```

Redimensionner le nombre de démons Hive LLAP

Utilisez la commande suivante pour réduire le nombre d'LLAP instances.

```
yarn app -flex llap0 -component llap -1
```

Pour plus d'informations, consultez [Ajustement du composant d'un service](#).

Chiffrement dans Hive

Cette section décrit les types de chiffrement pris EMR en charge par Amazon.

Chiffrement modulaire du parquet dans Hive

Le chiffrement modulaire Parquet fournit un contrôle d'accès et un chiffrement au niveau des colonnes afin d'améliorer la confidentialité et l'intégrité des données stockées au format de fichier Parquet. Cette fonctionnalité est disponible dans Amazon EMR Hive à partir de la version 6.6.0.

Les solutions de sécurité et d'intégrité précédemment prises en charge, notamment le chiffrement de fichiers ou le chiffrement de la couche de stockage, sont décrites dans la section [Options de chiffrement](#) du guide de EMR gestion Amazon. Ces solutions peuvent être utilisées pour les fichiers Parquet, mais l'exploitation des nouvelles fonctionnalités du mécanisme de cryptage intégré de Parquet permet un accès granulaire au niveau des colonnes, ainsi que des améliorations en termes de performances et de sécurité. Pour en savoir plus sur cette fonctionnalité, consultez la page github d'Apache intitulée [Parquet Modular Encryption](#).

Les utilisateurs transmettent les configurations aux lecteurs et rédacteurs Parquet à l'aide des configurations Hadoop. Les configurations détaillées permettant aux utilisateurs de configurer les lecteurs et les enregistreurs afin d'activer le chiffrement et d'activer les fonctionnalités avancées sont documentées à la page [PARQUET-1854 : Interface pilotée par les propriétés pour la gestion du chiffrement des parquets](#)

Exemples d'utilisation :

L'exemple suivant couvre la création et l'écriture dans une table Hive en utilisant AWS KMS pour gérer les clés de chiffrement.

1. Implémentez un KmsClient pour le AWS KMS service comme décrit dans le document [PARQUET-1373 : Outils de gestion des clés de chiffrement](#). L'exemple suivant montre un extrait de code d'implémentation.

```
package org.apache.parquet.crypto.keytools;

import com.amazonaws.AmazonClientException;
import com.amazonaws.AmazonServiceException;
import com.amazonaws.regions.Regions;
import com.amazonaws.services.kms.AWSKMS;
import com.amazonaws.services.kms.AWSKMSClientBuilder;
import com.amazonaws.services.kms.model.DecryptRequest;
import com.amazonaws.services.kms.model.EncryptRequest;
import com.amazonaws.util.Base64;
import org.apache.hadoop.conf.Configuration;
import org.apache.parquet.crypto.KeyAccessDeniedException;
```



```
import org.apache.parquet.crypto.ParquetCryptoRuntimeException;
import org.apache.parquet.crypto.keytools.KmsClient;
import org.slf4j.Logger;
import org.slf4j.LoggerFactory;

import java.nio.ByteBuffer;
import java.nio.charset.Charset;
import java.nio.charset.StandardCharsets;

public class AwsKmsClient implements KmsClient {

    private static final AWSKMS AWSKMS_CLIENT = AWSKMSClientBuilder
        .standard()
        .withRegion(Regions.US_WEST_2)
        .build();
    public static final Logger LOG = LoggerFactory.getLogger(AwsKmsClient.class);

    private String kmsToken;
    private Configuration hadoopConfiguration;

    @Override
    public void initialize(Configuration configuration, String kmsInstanceID, String
kmsInstanceURL, String accessToken) throws KeyAccessDeniedException {
        hadoopConfiguration = configuration;
        kmsToken = accessToken;
    }

    @Override
    public String wrapKey(byte[] keyBytes, String masterKeyIdentifier) throws
KeyAccessDeniedException {
        String value = null;
        try {
            ByteBuffer plaintext = ByteBuffer.wrap(keyBytes);

            EncryptRequest req = new
EncryptRequest().withKeyId(masterKeyIdentifier).withPlaintext(plaintext);
            ByteBuffer ciphertext = AWSKMS_CLIENT.encrypt(req).getCiphertextBlob();

            byte[] base64EncodedValue = Base64.encode(ciphertext.array());
            value = new String(base64EncodedValue, Charset.forName("UTF-8"));
        } catch (AmazonClientException ae) {
            throw new KeyAccessDeniedException(ae.getMessage());
        }
    }
}
```

```

        return value;
    }

    @Override
    public byte[] unwrapKey(String wrappedKey, String masterKeyIdentifier) throws
    KeyAccessDeniedException {
        byte[] arr = null;
        try {
            ByteBuffer ciphertext =
            ByteBuffer.wrap(Base64.decode(wrappedKey.getBytes(StandardCharsets.UTF_8)));
            DecryptRequest request = new
            DecryptRequest().withKeyId(masterKeyIdentifier).withCiphertextBlob(ciphertext);
            ByteBuffer decipheredtext =
            AWSKMS_CLIENT.decrypt(request).getPlaintext();
            arr = new byte[decipheredtext.remaining()];
            decipheredtext.get(arr);
        } catch (AmazonClientException ae) {
            throw new KeyAccessDeniedException(ae.getMessage());
        }
        return arr;
    }
}

```

2. Créez vos clés de AWS KMS chiffrement pour le pied de page ainsi que pour les colonnes auxquelles vos IAM rôles ont accès, comme décrit dans la section [Création de clés](#) dans le guide du AWS Key Management Service développeur. Le IAM rôle par défaut est EMR _ ECS _default.
3. Dans l'application Hive sur un EMR cluster Amazon, ajoutez le client ci-dessus à l'aide de l'ADD JARinstruction, comme décrit dans la documentation [Apache Hive Resources](#). Voici un exemple de déclaration :

```
ADD JAR 's3://location-to-custom-jar';
```

Une autre méthode consiste à ajouter le JAR à Hive à auxlib l'aide d'une action bootstrap. Voici un exemple de ligne à ajouter à l'action bootstrap :

```
aws s3 cp 's3://location-to-custom-jar' /usr/lib/hive/auxlib
```

4. Définissez les configurations suivantes :

```
set
  parquet.crypto.factory.class=org.apache.parquet.crypto.keytools.PropertiesDrivenCryptoFactory;
set
  parquet.encryption.kms.client.class=org.apache.parquet.crypto.keytools.AwsKmsClient;
```

5. Créez une table Hive au format Parquet, spécifiez les AWS KMS clés SERDEPROPERTIES et insérez-y des données :

```
CREATE TABLE my_table(name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe'
WITH SERDEPROPERTIES (
  'parquet.encryption.column.key'=<aws-kms-key-id-for-column-1>: credit_card',
  'parquet.encryption.footer.key'='<aws-kms-key-id-for-footer>')
STORED AS parquet
LOCATION "s3://<bucket>/<warehouse-location>/my_table";

INSERT INTO my_table SELECT
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',5) as
  name,
java_method ('org.apache.commons.lang.RandomStringUtils','randomAlphabetic',10) as
  credit_card
from (select 1) x lateral view posexplode(split(space(100),' ')) pe as i,x;

select * from my_table;
```

6. Vérifiez que lorsque vous créez une table externe au même endroit sans accès aux AWS KMS clés (par exemple, accès au IAM rôle refusé), vous ne pouvez pas lire les données.

```
CREATE EXTERNAL TABLE ext_table (name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe'
STORED AS parquet
LOCATION "s3://<bucket>/<warehouse-location>/my_table";

SELECT * FROM ext_table;
```

7. La dernière instruction doit générer l'exception suivante :

```
Failed with exception
java.io.IOException:org.apache.parquet.crypto.KeyAccessDeniedException: Footer key:
access denied
```

Chiffrement en transit en 2 HiveServer

À partir de la EMR version 6.9.0 d'Amazon, HiveServer 2 (HS2) est SSL activé/TLS/dans le cadre de la configuration de [Chiffrement en transit en 2 HiveServer](#) sécurité. Cela affecte la façon dont vous vous connectez à l'HS2 exécution sur un EMR cluster Amazon avec le chiffrement en transit activé. Pour vous connecter à HS2, vous devez modifier les valeurs TRUSTSTORE_PATH des TRUSTSTORE_PASSWORD paramètres et dans le JDBCURL. URL Voici un exemple de JDBC connexion pour HS2 avec les paramètres requis :

```
jdbc:hive2://HOST_NAME:10000/  
default;ssl=true;sslTrustStore=TRUSTSTORE_PATH;trustStorePassword=TRUSTSTORE_PASSWORD
```

Utilisez les instructions appropriées pour le chiffrement sur cluster 2 ou hors cluster 2 ci-dessous HiveServer.

On-cluster HS2 access

Si vous accédez à HiveServer 2 à l'aide du client Beeline après avoir accédé SSH au nœud principal, faites référence `/etc/hadoop/conf/ssl-server.xml` pour trouver les valeurs des TRUSTSTORE_PASSWORD paramètres TRUSTSTORE_PATH et à l'aide de la configuration `ssl.server.truststore.location` et `ssl.server.truststore.password`.

Les exemples de commandes suivants peuvent vous aider à récupérer ces configurations :

```
TRUSTSTORE_PATH=$(sed -n '/ssl.server.truststore.location/,+2p' /etc/hadoop/conf/  
ssl-server.xml | awk -F "[><]" '/value/{print $3}')  
TRUSTSTORE_PASSWORD=$(sed -n '/ssl.server.truststore.password/,+2p' /etc/hadoop/  
conf/ssl-server.xml | awk -F "[><]" '/value/{print $3}')
```

Off-cluster HS2 access

Si vous accédez à HiveServer 2 depuis un client extérieur au EMR cluster Amazon, vous pouvez utiliser l'une des approches suivantes pour obtenir le TRUSTSTORE_PATH et : TRUSTSTORE_PASSWORD

- Convertissez le PEM fichier créé lors [de la configuration de sécurité](#) en JKS fichier et utilisez-le lors de la JDBC connexion URL. Par exemple, avec openssl et keytool, utilisez les commandes suivantes :

```
openssl pkcs12 -export -in trustedCertificates.pem -inkey privateKey.pem -out
trustedCertificates.p12 -name "certificate"
keytool -importkeystore -srckeystore trustedCertificates.p12 -srcstoretype pkcs12
-destkeystore trustedCertificates.jks
```

- Vous pouvez également faire référence à `/etc/hadoop/conf/ssl-server.xml` pour trouver les valeurs des paramètres `TRUSTSTORE_PATH` et `TRUSTSTORE_PASSWORD` à l'aide de la configuration `ssl.server.truststore.location` et `ssl.server.truststore.password`. Téléchargez le fichier magasin de confiance sur la machine cliente et utilisez le chemin sur la machine cliente comme `TRUSTSTORE_PATH`.

Pour plus d'informations sur l'accès aux applications depuis un client extérieur au EMR cluster Amazon, consultez [Utiliser le JDBC pilote Hive](#).

Historique des versions de Hive

Le tableau suivant répertorie la version de Hive incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Hive

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-7.2.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client,

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
		hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-5.36.2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-7.1.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-7.0.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.15.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.14.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.13.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.12.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.11.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.11.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.10.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server
emr-6.10.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, tez-on-worker, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.9.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.9.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.8.1	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.8.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.7.0	3.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-5.36.1	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.36.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-6.6.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.35.0	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-6.5.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.4.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.3.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.3.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.2.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.2.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.1.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.1.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-6.0.1	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-6.0.0	3.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn, zookeeper-client, zookeeper-server
emr-5.34.0	2.3.8	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.33.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.33.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.32.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.32.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.31.1	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.31.0	2.3.7	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.30.2	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
emr-5.30.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.30.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mariadb-server, tez-on-yarn
mr-5.29.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.28.1	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn
emr-5.28.0	2.3.6	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, hudi, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.27.1	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.27.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.26.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.25.0	2.3.5	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.24.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.24.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.23.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.23.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.22.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.21.2	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.21.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.21.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.20.1	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.20.0	2.3.4	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.19.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.19.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.18.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.18.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.17.2	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.17.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.17.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, emr-s3-select, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.16.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.16.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.15.1	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.15.0	2.3.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.14.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.14.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.14.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.13.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.13.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.12.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.12.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.12.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.12.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.11.4	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.3	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.11.2	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.11.1	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.11.0	2.3.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.10.1	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.10.0	2.3.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.9.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.9.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.3	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.8.2	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.8.1	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.8.0	2.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.7.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.7.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.6.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.6.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.4	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.5.3	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.5.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.5.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.4.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
emr-5.4.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.3.2	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.3.1	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.3.0	2.1.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.2.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.2.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.2.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.2.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.1.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.1.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.0.3	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.0.2	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-5.0.1	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-5.0.0	2.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
emr-4.9.6	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.9.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.9.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.9.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.9.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.5	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.8.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.8.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.8.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.8.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.4	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.7.3	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.2	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.7.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
emr-4.7.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.6.1	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.6.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.5.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.4.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.3.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.2.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Étiquette Amazon EMR Release	Version de Hive	Composants installés avec Hive
emr-4.1.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httptfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
emr-4.0.0	1.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-namenode, hadoop-httptfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Notes de mise à jour de Hive par version

Rubriques

- [Amazon EMR 7.1.0 - Notes de mise à jour de Hive](#)
- [Amazon EMR 6.15.0 - Notes de mise à jour de Hive](#)
- [Amazon EMR 6.14.0 - Notes de mise à jour de Hive](#)
- [Amazon EMR 6.13.0 - Notes de mise à jour de Hive](#)
- [Amazon EMR 6.12.0 - Notes de mise à jour de Hive](#)
- [Amazon EMR 6.11.0 - Notes de mise à jour de Hive](#)

- [Amazon EMR 6.10.0 - Notes de mise à jour de Hive](#)
- [Amazon EMR 6.9.0 - Notes de mise à jour de Hive](#)
- [Amazon EMR 6.8.0 - Notes de mise à jour de Hive](#)
- [Amazon EMR 6.7.0 - Notes de mise à jour de Hive](#)
- [Amazon EMR 6.6.0 - Notes de mise à jour de Hive](#)

Amazon EMR 7.1.0 - Notes de mise à jour de Hive

Amazon EMR 7.1.0 - Modifications apportées à Hive

Type	Description
Correctif de bogue.	HIVE-24381 — La saisie de texte compressé renvoie 0 ligne si l'option ignorer l'en-tête/le pied de page est inclus.
Correctif de bogue.	HIVE-24190 — LLAP : ShuffleHandler pourrait renvoyer DISK _ ERROR _ EXCEPTION conformément à TEZ -4233.
Correctif de bogue.	HIVE-23073 — Shade Netty.
Correctif de bogue.	HIVE-23073 — Shade Netty et passe à netty 4.1.48.Final.
Correctif de bogue.	HIVE-23148 — Le flux de clients externes Llap est interrompu à cause du netty shading.
Correctif de bogue.	HIVE-25180 — Améliore Netty.
Correctif de bogue.	HIVE-24524 — LLAP ShuffleHandler : passez à Netty4 et supprimez la dépendance Netty3 de Hive dans la mesure du possible.
Correctif de bogue.	HIVE-28000 — Hive QL : la clause « not in » donne des résultats incorrects lorsque la coercition de type ne peut pas avoir lieu.

Type	Description
Correctif de bogue.	HIVE-27993 — Netty4 ShuffleHandler doit utiliser 1 thread de boss.
Upgrade	Met à jour Netty vers la version 4.1.100.Final
Upgrade	Améliore Jetty vers la version 9.4.53.v20231009
Upgrade	Améliore Zookeeper vers la version 3.9.1

Amazon EMR 7.1.0 - Modifications apportées à Hive

- Amazon EMR 7.1 met à niveau Hive vers Netty 4.1.100.Final afin de résoudre les failles de sécurité de Netty3. Comme il hive-druid-handler dépend de netty3, Hive n'a pas le chemin de classe de Hive hive-druid-handler JAR dans Amazon 7.1. EMR Une prochaine EMR version d'Amazon l'inclura dans le classpath de Hive une fois que le gestionnaire Druid sera compatible avec les versions 4.1.100.Final ou ultérieures de Netty. Contactez l' AWS assistance si vous avez besoin hive-druid-handler JAR des EMR versions 7.1 ou supérieures d'Amazon.

Amazon EMR 6.15.0 - Notes de mise à jour de Hive

Amazon EMR 6.15.0 - Changements liés à Hive

Type	Description
Fonctionnalité	Support pour TEZ-4397 — Pour l'ouverture fractionnée asynchrone de Tez, Hive prend désormais en charge les configurations Tez décrites dans. Ouverture fractionnée asynchrone Tez
Correctif de bogue	HIVE-25400 — Déplace la mise à jour du décalage vers. BytesColumnVector setValPreallocated

Type	Description
Correctif de bogue	HIVE-25190 — Corrigez de nombreuses petites allocations dans. BytesColumnVector
Upgrade	Permet de mettre à niveau Apache Hadoop vers la version 3.3.6.
Upgrade	HIVE-26684 — Mise à niveau vers maven-sha de-plugin la version 3.4.1.
Amélioration	Pour réduire le temps de démarrage du EMR cluster Amazon, supprimez 15 secondes de temps de veille du script de HCatalog démarrage.

Amazon EMR 6.14.0 - Notes de mise à jour de Hive

Amazon EMR 6.14.0 - Changements liés à Hive

Type	Description
Amélioration	HIVE-26762 : Supprimer l'élagage des opérandes HiveFilterSetOpTransposeRule
Correctif de bogue	HIVE-27582 : Ne pas mettre en cache le format d'entrée de la HBase table dans FetchOperator
Correctif de bogue	HIVE-26452 : NPE lors de la conversion JOIN vers une MAPJOIN JOIN colonne référencée plusieurs fois
Correctif de bogue	HIVE-26416 : AcidUtils.isRawFormatFile() lance InvalidProtocolBufferException un fichier non-ORC

Type	Description
Correctif de bogue	HIVE-26105 : Afficher les colonnes affiche des valeurs supplémentaires si la colonne Commentaires contient un caractère chinois spécifique
Correctif de bogue	HIVE-25864 : L'optimisation des requêtes Hive crée un mauvais plan pour le transfert des prédicats avec fonction de fenêtrage
Correctif de bogue	HIVE-25224 : Plusieurs INSERT instructions impliquant des tables avec des bucketing _versions résultats d'erreur différents
Correctif de bogue	HIVE-24151 : MultiDelimiter déplace les données si les chaînes contiennent des caractères non-ASCII
Correctif de bogue	HIVE-23606 : (LLAP) Retard dans le nettoyage pour DirectByteBuffer EncodedReaderImpl
Correctif de bogue	HIVE-22165 : La synchronisation introduite par HIVE-14296 SessionManager.closeSession entraîne une latence élevée sur un serveur Hive très occupé
Correctif de bogue	HIVE-21304 : Rendre l'utilisation de la version de compartiments plus robuste

Amazon EMR 6.13.0 - Notes de mise à jour de Hive

Amazon EMR 6.13.0 - Changements liés à Hive

Type	Description
Amélioration	Mettre à niveau les scripts Python pour qu'ils prennent en charge Python3
Amélioration	HIVE-27097 : Améliore la stratégie de nouvelle tentative pour MetaStore le client et le serveur
Correctif de bogue.	HIVE-21778 : CBO : « La structure n'est pas nulle » est considérée comme nullable, ce qui entraîne toujours un échec du filtre dans la requête
Correctif de bogue.	HIVE-21009 : Ajout de la possibilité pour l'utilisateur de définir un utilisateur de liaison
Correctif de bogue.	HIVE-22661 : Le compactage échoue sur une table non compartimentée avec des données chargées dans le chemin
Correctif de bogue.	HIVE-19718 : L'ajout de partitions en masse permet également de récupérer le tableau pour chaque partition
Correctif de bogue.	HIVE-22173 : La requête avec plusieurs vues latérales se bloque lors de la compilation
Correctif de bogue.	HIVE-27088 : Résultats incorrects lorsque les jointures internes et externes avec des filtres de post-jointure sont fusionnées
Correctif de bogue.	HIVE-21935 : Vectorisation des ruches : performance dégradée avec la vectorisation UDF

Type	Description
Correctif de bogue.	HIVE-25299 : La conversion de l'horodatage en types de données numériques est incorrecte pour les fuseaux autres que les fuseaux horaires UTC
Correctif de bogue.	HIVE-24626 : LLAP : les fils de lecture peuvent être privés si tous les threads de l'ascenseur d'E/S sont occupés pour être mis en file d'attente à un autre lecteur avec une file d'attente complète
Correctif de bogue.	HIVE-27029 : la requête hive échoue avec une erreur de fermeture du système de fichiers, refonte effectuée pour -26352 HIVE
Correctif de bogue.	HIVE-26352 : La vérification de l'accès à la file d'attente Tez échoue avec une GSS exception lors du compactage
Correctif de bogue.	HIVE-24590 : La journalisation des opérations divulgue toujours les appendices log4j
Correctif de bogue.	HIVE-24552 : Possibilité de fuite ou d'accumulation de HMS raccords loadDynamicPartitions
Correctif de bogue.	HIVE-27069 : Résultats incorrects avec la jointure par bucket map
Correctif de bogue.	HIVE-27344 : Ajouter une vérification nulle dans #close RecordReaderImpl
Correctif de bogue.	HIVE-27439 : Support de l'espace en décimal
Correctif de bogue.	HIVE-27267 : Résultats incorrects lors de la jointure par bucket map sur une colonne décimale avec sous-requête

Type	Description
Correctif de bogue.	HIVE-21986 : Interface utilisateur HiveServer Web : définition du paramètre Strict-Transport-Security dans l'en-tête de réponse par défaut
Correctif de bogue.	HIVE-22148 : Les jetons de délégation S3A ne sont pas ajoutés dans la configuration des tâches du compacteur.
Correctif de bogue.	HIVE-22622 : Hive permet de créer une structure avec des noms d'attributs dupliqués
Correctif de bogue.	HIVE-22008 : LIKE L'opérateur doit faire correspondre l'entrée multiligne
Correctif de bogue.	HIVE-23144 LLAP : Laisse QueryTracker le nettoyage se poursuivre serviceStop
Correctif de bogue.	HIVE-22391 : NPE lors de la vérification du cache des résultats des requêtes Hive
Correctif de bogue.	HIVE-23305 : NullPointerException en LlapTaskSchedulerService addNode raison des conditions de course
Correctif de bogue.	HIVE-22178 : Parquet jeté après FilterPredicate CastException SchemaEvolution
Correctif de bogue.	HIVE-21517 : Corriger AggregateStatsCache
Correctif de bogue.	HIVE-21825 : Améliore le message d'erreur du client lorsque la haute disponibilité active/passive est activée
Correctif de bogue.	HIVE-23389 : FilterMergeRule peut conduire à AssertionError

Type	Description
Correctif de bogue.	HIVE-22767 : Beeline n'analyse pas correctement les points-virgules dans les commentaires
Correctif de bogue.	HIVE-22996 : l' BasicStats analyse doit vérifier de manière proactive la présence d'une chaîne nulle ou vide
Correctif de bogue.	HIVE-22808 : HiveRelFieldTrimmer ne gère pas HiveTableFunctionScan
Correctif de bogue.	HIVE-22437 : Cache de LLAP métadonnées lors du verrouillage NPE des métadonnées.
Correctif de bogue.	HIVE-22606 : AvroSerde enregistre avro.schema.literal sous le niveau INFO
Correctif de bogue.	HIVE-22713 : La propagation constante ne doit pas être effectuée pour la structure Join-Fil (*) - RS
Correctif de bogue.	HIVE-21624 : LLAP : Les métriques du processeur au niveau du thread sont cassées
Correctif de bogue.	HIVE-22815 : réduit la création inutile d'objets de système de fichiers dans MROutput
Correctif de bogue.	HIVE-23060 : Échec de la requête avec erreur « Le regroupement des ensembles d'expressions n'est pas dans la touche GROUP BY. Une erreur s'est produite à proximité du jeton »
Correctif de bogue.	HIVE-22236 : Impossible de créer une vue en sélectionnant la vue contenant une sous-requête NOT IN

Type	Description
Correctif de bogue.	HIVE-19886 : Les journaux peuvent être dirigés vers 2 fichiers si <code>—hiveconf hive.log.file</code> est utilisé
Correctif de bogue.	HIVE-20620 : collisions manifestes lors de l'insertion dans des tables MM triées par compartiments avec partitionnement dynamique
Correctif de bogue.	HIVE-14557 : Nullpointer lorsque Mapjoin et Mapjoin sont activés SkewJoin
Correctif de bogue.	HIVE-20471 : problèmes pour obtenir le chemin de base de données par défaut
Correctif de bogue.	HIVE-20598 : Corriger les fautes de frappe dans les calculs HiveAlgorithmsUtil
Correctif de bogue.	HIVE-14737 : Problème lors de l'accès à <code>/logs</code> dans une interface utilisateur Web Kerberized Hive Server 2
Correctif de bogue.	HIVE-20733 : <code>GenericUDFOPEqual NS</code> ne peut pas utiliser <code>=</code> dans les descriptions des plans
Correctif de bogue.	HIVE-20848 : Après la configuration, la <code>UpdateInputAccessTimeHook</code> requête échoue avec <code>Table Not Found</code> .
Correctif de bogue.	HIVE-18929 : La méthode <code>humanReadableInt</code> du <code>HiveStringUtils</code> fichier <code>.java</code> comporte une condition de race.
Correctif de bogue.	HIVE-20841 : LLAP : Rendre les ports dynamiques configurables

Type	Description
Correctif de bogue.	HIVE-20930 : le FILTER mode VectorCoalesce en cours ne prend pas effet
Correctif de bogue.	HIVE-21007 : Semi-join + Union peut mener à de mauvais plans
Correctif de bogue.	HIVE-21074 : L'élagage des requêtes de table à compartiments Hive ne fonctionne pas pour l'état du SI NOT NULL
Correctif de bogue.	HIVE-21223 : CachedStore renvoie une partition nulle lorsque la partition n'existe pas
Correctif de bogue.	HIVE-19625 : Exception potentielle NPE et masquage réelle dans Hive# copyFiles
Correctif de bogue.	HIVE-17020 : Un dédup RS agressif peut supprimer de manière incorrecte une branche d'arbre OP
Correctif de bogue.	HIVE-20168 : Journalisation cachée ReduceSinkOperator
Correctif de bogue.	HIVE-20879 : L'utilisation de null dans une expression de projection entraîne CastException
Correctif de bogue.	HIVE-20888 TxnHandler : sort () appelé sur des listes immuables
Correctif de bogue.	HIVE-19948 : HiveCli ne divise pas correctement la commande par point-virgule si des guillemets sont à l'intérieur de la chaîne
Correctif de bogue.	HIVE-20621 : GetOperationStatus appelé dans resultset.next provoquant une lenteur incrémentielle

Type	Description
Correctif de bogue.	HIVE-20854 : Valeurs par défaut sensibles : l'intervalle de battement de cœur Zookeeper de Hive est de 20 minutes, passez à 2
Correctif de bogue.	HIVE-20330 : HCatLoader impossible de gérer plusieurs InputJobInfo objets pour une tâche comportant plusieurs entrées
Correctif de bogue.	HIVE-20787 : le MapJoinBytesTableContainer dummyRow boîtier ne gère pas la réutilisation
Correctif de bogue.	HIVE-20331 : La requête avec union tout, vue latérale et jointure échoue avec « Impossible de trouver le parent dans l'opérateur enfant »
Correctif de bogue.	HIVE-19968 : UDF l'exception n'est pas supprimée
Correctif de bogue.	HIVE-20410 : l'annulation de l'insertion et de la réécriture sur la table transactionnelle provoque l'erreur « Historique insuffisant disponible pour... »
Correctif de bogue.	HIVE-20059 : Le streaming Hive devrait essayer le préfixe shade sans condition, sauf exception
Correctif de bogue.	HIVE-19424 : Entrée NPE MetadataFormatters
Correctif de bogue.	HIVE-20355 : Paramètre de nettoyage de HiveConnection setSchema
Correctif de bogue.	HIVE-20858 : Le sérialiseur n'est pas correctement initialisé avec la configuration dans Utilities. createEmptyBuckets

Type	Description
Correctif de bogue.	HIVE-20424 : Schematool ne doit pas polluer l'historique de Beeline
Correctif de bogue.	HIVE-20338 : LLAP : Forcer un identifiant de fichier synthétique pour les systèmes de fichiers qui ont des impls de protocole avec une sémantique de mutation HDFS POSIX
Correctif de bogue.	HIVE-11708 : Les opérateurs logiques augmentent avec ClassCastExceptions NULL
Correctif de bogue.	HIVE-21082 : DansHPL/SQL, l'instruction declare ne prend pas en charge les variables de type caractère
Correctif de bogue.	HIVE-16690 : Configuration de la périphérie cartésienne du produit Tez en fonction de la taille du cluster LLAP
Correctif de bogue.	HIVE-21296 : Suppression d'une exception de lancement de partition varchar
Correctif de bogue.	HIVE-14516 : OrcInputFormat SplitGenerator. callInternal
Correctif de bogue.	HIVE-20981 : streaming/ fuites AbstractRecordWriter HeapMemoryMonitor
Correctif de bogue.	HIVE-20043 : HiveServer 2 : SessionState possède un bloc de synchronisation statique autour d'un AtomicBoolean
Correctif de bogue.	HIVE-20191 : l'application du PreCommit correctif n'échoue pas si le correctif est vide

Type	Description
Correctif de bogue.	HIVE-20400 : la création d'une table doit toujours utiliser un chemin entièrement qualifié pour éviter toute ambiguïté potentielle avec FS
Correctif de bogue.	Ajouter une vérification nulle skewedInfo avant d'accéder aux colonnes asymétriques

Amazon EMR 6.12.0 - Notes de mise à jour de Hive

Amazon EMR 6.12.0 - Changements liés à Hive

Type	Description
Amélioration	Support ajouté pour les JDK environnements d'exécution 11 et JDK 17
Amélioration	Ajoutez la possibilité de demander des noms de colonnes de mots clés réservés et sensibles aux majuscules lors de l'utilisation de S3 Select. Pour l'utiliser, définissez la propriété de la table au format « s3select.column.mapping » = " <i>column1:fieldName1</i> , <i>column2:fieldName2</i> , ..."
Amélioration	HIVE-23133 : Les opérations numériques peuvent avoir des résultats différents selon les arches matérielles
Amélioration	HIVE-27145 : À utiliser StrictMath pour les fonctions mathématiques restantes comme suite à -23133 HIVE
Correctif de bogue.	Correction de l'incompatibilité des caractères génériques dans get_partitions_by_filter et get_num_partitions_by_filter causée par le

Type	Description
	portage -22900 dans Hive 6.4.0 HMS APIs HIVE EMR
Correctif de bogue.	HIVE-26736 : Échec de l'autorisation pour les vues imbriquées comportant une clause WITH
Correctif de bogue.	HIVE-22416 : Les journaux des opérations liés à la résonance magnétique sont manquants lorsque l'exécution parallèle est activée
Correctif de bogue.	HIVE-19653 : Pushdown de prédicat incorrect pour groupby avec des ensembles de regroupement
Correctif de bogue.	HIVE-22094 : échec des requêtes avec : ClassCastException hive.ql.exec.vector. DecimalColumnVectorne peut pas être converti en Hive.QL.Exec.Vector.Decimal64 ColumnVec tor
Correctif de bogue.	HIVE-26340 : L'PTFopérateur vectorisé échoue si la requête comporte une fonction de fenêtre en majuscules
Correctif de bogue.	HIVE-26184 : COLLECT _ SET avec GROUP BY est très lent lorsque certaines touches sont très asymétriques
Correctif de bogue.	HIVE-26373 : ClassCastException lors de la lecture des horodatages d'une table contenant des données HBase Avro

Type	Description
Correctif de bogue.	HIVE-26388 : ClassCastException lorsqu'il y a une colonne autre que de type chaîne dans la table source de la CTAS requête Upgrade HIVE-26172 : Hive - Upgrade Ant vers la version 1.10.11 en raison de -2021-36373 et -2021-36374 CVE CVE
Correctif de bogue.	HIVE-26114 : Corriger la connexion jdbc hiveserver2 en utilisant la commande dfs avec un espace de préfixe provoquera une exception
Correctif de bogue.	HIVE-26396 : La fonction trunc a un problème d'interception de précision et le résultat a de nombreux 0
Correctif de bogue.	HIVE-26446 : <u>Impossible</u> de renseigner le HiveProtoLoggingHook TablesWritten champ pour les tables partitionnées.
Correctif de bogue.	HIVE-26639 : ConstantVectorExpression et ExplainTask ne doit pas s'appuyer sur le jeu de caractères par défaut
Correctif de bogue.	HIVE-22670 : ArrayIndexOutOfBoundsException lorsqu'un lecteur vectorisé est utilisé pour lire un dossier de parquet
Correctif de bogue.	HIVE-23607 : Problème d'autorisation : la création d'une vue sur une autre vue réussit mais la modification de la vue échoue
Correctif de bogue.	HIVE-25498 : Une requête contenant plus de 31 fonctions distinctes renvoie un résultat erroné
Correctif de bogue.	HIVE-25780 : DistinctExpansion crée plus de 64 ensembles de regroupement II

Type	Description
Correctif de bogue.	HIVE-23868 : Spécification de la fonction de fenêtrage : support 0 précédant/suivant
Correctif de bogue.	HIVE-24539 : la génération du OrclnputFormat schéma doit respecter le délimiteur de colonne
Correctif de bogue.	HIVE-23476 LLAP : Préallouez également des arènes pour le cas de carte
Correctif de bogue.	HIVE-25806 : Fuite possible LlapCache AwareFs - Parquet, IO LLAP
Correctif de bogue.	HIVE-23498 : Désactiver la méthode HTTP Trace sur ThriftHttpCliService
Correctif de bogue.	HIVE-25729 : ThriftUnionObjectInspector doit être notifié lorsqu'il est complètement invité
Correctif de bogue.	HIVE-23846 : Évitez la sérialisation et la désérialisation inutiles des vecteurs de bits
Correctif de bogue.	HIVE-24233 : sauf que la sous-requête lance un pointeur nul avec cbo désactivé
Correctif de bogue.	HIVE-24276 : Vulnérabilité liée à HiveServer 2 scripts intersites dans loggerconf.jsp () XSS
Correctif de bogue.	HIVE-25721 : Le résultat de la jointure externe est erroné
Correctif de bogue.	HIVE-25223 : Sélectionner avec limite ne renvoie aucune ligne sur une table non native
Correctif de bogue.	HIVE-25794 CombineHiveRecordReader : les instructions de log dans une boucle entraînent une pression sur la mémoire

Type	Description
Correctif de bogue.	HIVE-23602 : Utiliser le package Java simultané pour le set de descripteurs d'opérations
Correctif de bogue.	HIVE-24045 : Aucune journalisation liée à la création de la base de données par défaut
Correctif de bogue.	HIVE-24305 : le schéma décimal avro ne remplit pas correctement l'échelle/la précision si la valeur est entre guillemets
Correctif de bogue.	HIVE-25844 : Des erreurs de désérialisation exceptionnelles peuvent entraîner la fermeture immédiate de Beeline
Correctif de bogue.	HIVE-25040 : Supprimer la cascade de base de données ne peut pas supprimer les fonctions persistantes
Correctif de bogue.	HIVE-23501 : VectorDeserializeRow lorsque AOOB les types complexes sont convertis en types primitifs
Correctif de bogue.	HIVE-23704 : Le HTTP serveur Thrift ne gère pas correctement le gestionnaire d'authentification
Correctif de bogue.	HIVE-23529 : CTAS est cassé pour uniontype lorsque row_deserialize
Correctif de bogue.	HIVE-24144 : La getIdentifierQuote chaîne HiveDatabaseMetaData renvoie une valeur incorrecte
Correctif de bogue.	HIVE-23850 : Autoriser PPD lorsque le sujet n'est pas une colonne avec des ensembles de regroupement présents

Type	Description
Correctif de bogue.	HIVE-24036 : Exception Kryo lors de la sérialisation du plan d'appel getSplits UDF
Correctif de bogue.	HIVE-25919 : ClassCastException lorsque vous insérez un prédicat de colonne booléen HBaseStorageHandler
Correctif de bogue.	HIVE-25261 : RetryingHMSHandler doit envelopper le MetaException avec une brève description de la cible
Correctif de bogue.	HIVE-24792 : Fuite de fil potentielle en cours de fonctionnement
Correctif de bogue.	HIVE-23409 : Si la réouverture de TezSession l'application échoue en raison de la panne du service Timeline, le formulaire par défaut SessionPool est fermé TezSession après une nouvelle tentative
Correctif de bogue.	HIVE-23615 : Ne défiez pas les pointeurs nuls dans la classe Beeline Commands
Correctif de bogue.	HIVE-24849 : crée un délai d'expiration du socket de table externe lorsque l'emplacement contient un grand nombre de fichiers (affecte la version 3.1.2)
Correctif de bogue.	HIVE-24193 : La requête de sélection sur la table Hive Acid renommée ne produit aucune sortie
Correctif de bogue.	HIVE-25209 : SELECT requête avec SUM fonction produisant un résultat inattendu

Type	Description
Correctif de bogue.	HIVE-23666 : checkHashMode L'efficacité est ignorée lorsqu'un opérateur groupby n'a pas d'ensemble de regroupement
Correctif de bogue.	HIVE-23873 : L'interrogation de la JDBCStorageHandler table Hive échoue lorsque cette option est désactivée NPE CBO
Correctif de bogue.	HIVE-24149 : HiveStreamingConnection ne ferme pas la connexion HMS
Correctif de bogue.	HIVE-25561 : La tâche supprimée ne doit pas valider le fichier. (affecte les versions 2.x et 3.x)
Correctif de bogue.	HIVE-25683 : Fermez le lecteur. AcidUtils isRawFormatDossier
Correctif de bogue.	HIVE-24294 : les TezSessionPool sessions peuvent être lancées AssertionError
Correctif de bogue.	HIVE-24182 : Délivrance d'une autorisation de ranger permanente UDFs
Correctif de bogue.	HIVE-22805 : La vectorisation avec un tableau conditionnel ou une carte n'est pas implémentée et génère une erreur
Correctif de bogue.	HIVE-22828 : Decimal64 : NVL et les CASE instructions convertissent implicitement decimal64 en 128
Correctif de bogue.	HIVE-21398 : Les colonnes contenant des statistiques estimées ne doivent pas être considérées comme des clés uniques

Type	Description
Correctif de bogue.	HIVE-22490 : L'ajout de fichiers jar contenant des caractères spéciaux dans leur chemin génère une erreur
Correctif de bogue.	HIVE-22700 : Les compactages peuvent entraîner une fuite de mémoire lorsqu'ils ne sont pas autorisés
Correctif de bogue.	HIVE-22053 : Le nom de la fonction n'est pas normalisé lors de la création de la fonction
Correctif de bogue.	HIVE-22595 : Les insertions de partitions dynamiques échouent sur la table Avro avec un schéma externe
Correctif de bogue.	HIVE-21795 : La ligne récapitulative du cumul peut être manquante lorsqu'un mapjoin se produit sur une table partitionnée
Correctif de bogue.	HIVE-22987 : ClassCastException dans VectorCoalesce quand est nul DataTypePhysicalVariation
Correctif de bogue.	HIVE-22814 : ArrayIndexOutOfBounds dans la vectorisation getData Type PhysicalVariation
Correctif de bogue.	HIVE-22523 : Le gestionnaire d'erreurs LlapRecordReader peut se bloquer si sa file d'attente est pleine
Correctif de bogue.	HIVE-21796 : ArrayWritableObjectInspector.equals peut prendre du temps à O (2^nesting_depth)
Correctif de bogue.	HIVE-22929 : Performance : l'analyse des identifiants entre guillemets utilise Regex.jetable via String.replaceAll()

Type	Description
Correctif de bogue.	HIVE-21641 : Le client externe Llap renvoie des colonnes décimales avec une précision/échelle différente de celle de Beeline
Correctif de bogue.	HIVE-22207 : Tez : se SplitGenerator lance NumberFormatException lorsque « dfs.block.size » sur le cluster est de « 128 m »
Correctif de bogue.	HIVE-22114 : La requête d'insertion pour une table à insertion uniquement partitionnée échoue lorsque tous les compartiments sont vides
Correctif de bogue.	HIVE-22165 : Synchronisation introduite par HIVE -14296 le. SessionManager closeSession provoque une latence élevée dans un serveur Hive très fréquenté
Correctif de bogue.	HIVE-22744 : car TezTask le sommet avec plusieurs arêtes de sortie doit avoir une mémoire de tri proportionnelle
Correctif de bogue.	HIVE-22072 : La modification de la table pour modifier une colonne ne met pas à jour les références aux contraintes
Correctif de bogue.	HIVE-22075 : Corrige la régression max-reducers=1 à partir de -14200 HIVE
Correctif de bogue.	HIVE-22527 : Hive on Tez : le travail de fusion de petits fichiers sera soumis dans une autre file d'attente (file d'attente par défaut)
Correctif de bogue.	HIVE-22816 : QueryCache : Les requêtes utilisant des vues peuvent être mises en cache après l'expansion CTE

Type	Description
Correctif de bogue.	HIVE-22733 : Après avoir désactivé la propriété du journal des opérations dans Hive, HS2 enregistrement du journal des opérations
Correctif de bogue.	HIVE-22699 : Le masque UDFs doit masquer la valeur numérique 0
Correctif de bogue.	HIVE-23356 : L'agrégation de hachage est toujours désactivée lors du traitement des requêtes avec des ensembles de groupements d'expressions.
Correctif de bogue.	HIVE-21568 : HiveRelOptUtil isRowFilteringLe plan doit ignorer le projet
Correctif de bogue.	HIVE-21760 : L'optimisation du travail partagé doit être ignorée pour les jointures SMB
Correctif de bogue.	HIVE-22712 : Le ReExec pilote exécute et soumet la requête dans la file d'attente par défaut, quelle que soit la file d'attente définie par l'utilisateur
Correctif de bogue.	HIVE-21397 : BloomFilter pour Hive, la table Managed [ACID] ne fonctionne pas comme prévu
Correctif de bogue.	HIVE-23011 : L'optimiseur de travail partagé doit vérifier les prédicats résiduels lors de la comparaison des jointures
Correctif de bogue.	HIVE-21412 : PostExecOrcFileDump ne fonctionne pas avec les tables ACID

Type	Description
Correctif de bogue.	HIVE-22201 : ConvertJoinMapJoin # checkShuffleSize ForLargeTable lance ArrayIndexOutOfBoundsException si aucune grande table n'est sélectionnée
Correctif de bogue.	HIVE-21971 : HS2 fuite du chargeur de classe à cause de `ReflectionUtils: : CONSTRUCTOR_CACHE` avec des fonctions temporaires + Generic UDF
Correctif de bogue.	HIVE-21368 : Vectorisation : Decimal64 inutile - > conversion HiveDecimal
Correctif de bogue.	HIVE-25416 : fuite de mémoire dans le métastore Hive à cause d'un bogue datanucleus-api-jdo
Correctif de bogue.	HIVE-22219 : La désactivation d'un gestionnaire de nœuds bloque le redémarrage du service LLAP
Correctif de bogue.	HIVE-21793 : CBO récupère les statistiques des colonnes même si hive.stats.fetch.column.stats est défini sur false
Correctif de bogue.	HIVE-22163 : CBO : L'activation CBO active l'estimation des statistiques, même lorsque l'estimation est désactivée
Correctif de bogue.	HIVE-18735 : Crée une table comme si elle perdait un attribut transactionnel
Correctif de bogue.	HIVE-22433 : Hive JDBC Storage Handler : résultats incorrects extraits depuis et depuis la source de données BOOLEAN TIMESTAMP DataType JDBC

Type	Description
Correctif de bogue.	HIVE-19430 : ObjectStore cleanNotificationEvents OutOfMemory sur un grand nombre d'événements en attente
Correctif de bogue.	HIVE-20785 : Nom de clé incorrect dans le. JDBC DatabaseMetaData getPrimaryKeys méthode
Correctif de bogue.	HIVE-16116 : Beeline se lance NPE lorsque beeline.hiveconfvariables= {} dans beeline.properties
Correctif de bogue.	HIVE-20066 : hive.load.data.owner est comparé au principal complet
Correctif de bogue.	HIVE-20489 : Expliquer le plan des blocages de requêtes
Correctif de bogue.	HIVE-21033 : L'oubli de fermer l'opération coupe encore 2 sorties HiveServer
Correctif de bogue.	HIVE-19888 : Avertissement trompeur « METASTORE FILTER __ HOOK sera ignoré » de SessionState
Correctif de bogue.	HIVE-20303 : tableau de base de INSERT OVERWRITE TABLE données (...) PARTITION IF NOT EXISTS lance InvalidTableException
Correctif de bogue.	HIVE-16144 : CompactionInfo n'a pas d'égals/ hashCode mais est utilisé dans Set
Correctif de bogue.	HIVE-20818 : Les vues créées avec une WHERE sous-requête considéreront les vues référencées dans la sous-requête comme une entrée directe

Type	Description
Correctif de bogue.	HIVE-21005 LLAP : Lire plus de bandes par fuite ZlibCodecs
Correctif de bogue.	HIVE-20771 : LazyBinarySerDe échoue sur les structures vides.
Correctif de bogue.	HIVE-18852 : Message d'erreur trompeur lors de la validation de la table modifiée
Correctif de bogue.	HIVE-21124 :HPL/SQLne supporte pas la déclaration CREATE TABLE LIKE
Correctif de bogue.	HIVE-20935 : Le téléchargement de l'archive tar du package llap échoue, ce qui entraîne un échec du démarrage du service EC2 LLAP
Correctif de bogue.	HIVE-20409 : Hive ACID : la mise à jour/supprimer/fusionner ne nettoie pas le répertoire de mise en scène HDFS
Correctif de bogue.	HIVE-20570 : Le plan de l'union ALL avec hive.optimize.union.remove=true est incorrect
Correctif de bogue.	HIVE-20421 : Entité de caractère « \ b » illégale dans hive-default.xml.template
Correctif de bogue.	HIVE-19133 : Les mesures de performance par phase de HS2 WebUI ne s'affichent pas correctement
Correctif de bogue.	HIVE-18977 : La liste des partitions renvoie des résultats différents avec et directement JDO SQL
Correctif de bogue.	HIVE-20034 : Annulation des modifications apportées à la gestion des MetaStore exception s pour des raisons de rétrocompatibilité

Type	Description
Correctif de bogue.	HIVE-20672 : Le thread de connexion LlapTaskSchedulerService doit signaler chaque intervalle fixe
Correctif de bogue.	HIVE-12812 : Activez mapred.input.dir.recursive par défaut pour prendre en charge l'union avec la fonction d'agrégation
Correctif de bogue.	HIVE-20147 : L'ingestion du streaming Hive est satisfaite de la journalisation synchronisée
Correctif de bogue.	HIVE-19203 : Problème de sécurité des threads dans HiveMetaStore
Correctif de bogue.	HIVE-20091 : Tez : ajout d'informations de sécurité pour la sortie FileSinkOperator
Correctif de bogue.	HIVE-16906 : Hive ATSHook doit vérifier la présence de yarn.timeline-service.enabled avant de se connecter à ATS
Correctif de bogue.	HIVE-20714 : SHOW tblproperties pour une seule propriété renvoie la valeur dans la colonne name
Correctif de bogue.	HIVE-24730 : Les classes Shims remplacent silencieusement les valeurs de hive-site.xml et tez-site.xml
Correctif de bogue.	HIVE-22055 : le nombre sélectionné donne un résultat incorrect après le chargement des données depuis un fichier texte

Amazon EMR 6.11.0 - Notes de mise à jour de Hive

Amazon EMR 6.11.0 - Changements liés à Hive

Type	Description
Amélioration	Ajout de la prise en charge de la suppression multithread de partitions afin d'améliorer les performances de suppression de partitions
Amélioration	Support de lecture des fichiers de requête Hive codés
Amélioration	Gestionnaire Tez Shuffle activé par défaut pour les tâches Hive on Tez
Bogue	Ajout d'une option pour activer la distribution déterministe des clés aux réducteurs afin de corriger un résultat incorrect lorsque hive.groupby.skewindata est activé (signalé en -20220) HIVE
Bogue	Correction d'un échec du calcul des statistiques lorsque le nom de partition par défaut est configuré
Bogue	Respectez tous les paramètres de SSL personnalisés transmis lorsqu'ils SSL sont configurés prêts à l'emploi pour HiveServer 2 dans un cluster avec le chiffrement en transit activé
Rétroportage	HIVE-23617 : Problèmes d'API de stockage résolus FindBug
Rétroportage	HIVE-26408 : Vectorisation : corrigez la désallocation des colonnes à gratter, ne réutilisez pas un enfant comme sortie ConstantVectorExpression

Type	Description
Rétroportage	HIVE-23614 : Passez toujours à HiveConfig removeTempOr DuplicateFiles
Rétroportage	HIVE-23354 : Supprimer la vérification de la taille du fichier dans compareTempOr Duplicate Files
Rétroportage	HIVE-20344 : Corrigé PrivilegeSynchronizer pour SBA le lancer. AccessControlException La propriété hive.privilege.synchronize a également été introduite pour désactiver le synchronisateur de privilèges
Rétroportage	HIVE-15826 : Support de la configuration de « serialization.encoding » pour tous SerDes
Rétroportage	HIVE-18284 : Correction NPE lors de l'insertion de données avec la clause « distribute by » avec optimisation du tri Dynpart
Rétroportage	HIVE-24930 : Opérateur . setDone() le court-circuit provenant de l'opération enfant n'est pas utilisé dans le chemin de code vectorisé (si == 1) childSize
Rétroportage	HIVE-24523 : Le chemin de lecture vectorisé pour LazySimpleSerde ne respecte pas l'horodatage SERDEPROPERTIES
Rétroportage	HIVE-23265 : Les ensembles de lignes dupliqués sont renvoyés avec les paramètres Limit et Offset définis
Rétroportage	HIVE-21492 : VectorizedParquetRecordReader impossible de lire le fichier de parquet généré à l'aide d'un outil économique/personnalisé

Type	Description
Rétroportage	HIVE-22540 : Vectorisation : les colonnes Decimal64 ne fonctionnent pas avec. Vectorize dBatchUtil makeLikeColumnVecteur ()
Rétroportage	HIVE-22588 : Videz les lignes restantes pour le reste des ensembles de regroupement lorsque vous changez de mode de regroupement vectoriel
Rétroportage	HIVE-22551 : BytesColumnVector initBuffer doit nettoyer le vecteur et la longueur de manière cohérente
Rétroportage	HIVE-22448 : CBO : Élargissez le nombre multiple de manière distincte à l'aide d'une touche de regroupement
Rétroportage	HIVE-22248 : Corrige les problèmes persistants liés aux statistiques
Rétroportage	HIVE-22210 : La vectorisation peut réutiliser les colonnes de sortie de calcul impliquées dans le filtrage
Rétroportage	HIVE-21531 : Vectorisation : tous les NULL hashcodes ne sont pas calculés avec Murmur3
Rétroportage	HIVE-20419 : Vectorisation : empêche la mutation de VectorPartitionDesc après son utilisation dans une clé de hachage
Rétroportage	HIVE-19388 : ClassCastException lors de l'initialisation VectorMapJoinCommonOperator
Rétroportage	HIVE-21584 : Préparation à Java 11 : le chargeur de classes système n'est pas URLClassLoader

Type	Description
Rétroportage	HIVE-25107 : La journalisation des chemins de classe doit être au même niveau (#2271) DEBUG
Rétroportage	HIVE-22097 : java.util incompatible. ArrayList pour Java 11
Rétroportage	HIVE-23938 : LLAP : JDK11 - certains arguments jvm liés à la rotation des fichiers journaux GC ne peuvent plus être utilisés
Rétroportage	HIVE-26226 : Exclure jdk.tools dep de hive-metastore dans upgrade-acid
Rétroportage	HIVE-17879 : Mettre à jour le plugin Datanucleus Maven
Rétroportage	HIVE-27004 : DateTimeFormatterBuilder # appendZoneText ne peut pas analyser « UTC + » dans les versions Java supérieures à 8
Rétroportage	HIVE-16812 : VectorizedOrcAcidRowBatchReader ne filtre pas les événements de suppression
Rétroportage	HIVE-17917 : VectorizedOrcAcidRowBatchReader computeOffsetAndOptimisation des compartiments
Rétroportage	HIVE-19985 ACID : ignore le décodage des sections ROW __ID pour les requêtes en lecture seule
Rétroportage	HIVE-20635 : VectorizedOrcAcidRowBatchReader ne filtre pas les événements de suppression pour les fichiers originaux

Type	Description
Upgrade	Mettre à niveau Javadoc vers 3.3.1
Upgrade	Mettre à niveau Javassist vers la version 3.24.1-GA
Upgrade	Mise à jour apache-directory-server en 2.0.0-M14

Nouvelles configurations

Nom	Classification	Description
hive.metastore.fs.drop.partition.threads	hive-site	Nombre de threads principaux dans le pool de threads de la partition de dépôt.
hive.metastore.fs.drop.partition.keepalive.time	hive-site	Durée en secondes pendant laquelle un thread asynchrone de partition inactif (issu du pool de threads) attendra l'arrivée d'une nouvelle tâche avant de se terminer.
hive.metastore.fs.drop.partition.threadpool.max.queue.size	hive-site	Taille de file d'attente maximale à utiliser dans le pool de threads pour supprimer des partitions du système de fichiers.
hive.groupby.enable.deterministic.distribution	hive-site	Activez la distribution déterministe des clés aux réducteurs. Il transmettra une valeur initiale constante lors de l'appel de la fonction rand

Nom	Classification	Description
		utilisée pour le partitionnement aléatoire.
hive.privilege.synchronizer	hive-site	S'il faut synchroniser périodiquement les privilèges de l'autorisateur externe en HiveServer 2.
hive.cli.query.file.encoding	hive-site	Codage de fichier pour tous les types de fichiers de requête (fichier de requête, fichier de requête d'initialisation, fichier rc, etc.) fournis dans les arguments cli.
hive.emr.tez.shuffle.enabled	hive-site	Les jobs Hive on Tez utilisent désormais tez_shuffle par défaut au lieu de mapreduce_shuffle comme gestionnaire de shuffle par défaut.

Configurations obsolètes

Les propriétés de configuration suivantes sont obsolètes en raison de l'erreur [HIVE-23354](#) et ne sont plus prises en charge par les EMR versions 6.11.0 et supérieures d'Amazon.

Nom	Valeur par défaut
hive.mapred.reduce.tasks.speculative.execution	false
tez.am.speculation.enabled	false

Amazon EMR 6.10.0 - Notes de mise à jour de Hive

Amazon EMR 6.10.0 - Changements liés à Hive

Type	Description
Fonctionnalité	Activez les contrôles d'accès AWS Lake Formation basés pour les requêtes Apache Hive (écriture) via IAM Passthrough (Hive CLI / Steps) . API
Amélioration	Désactiver la configuration <code>hive.log.explain.output</code> par défaut pour réduire la taille du journal
Rétroportage	HIVE-26408 : Vectorisation : corrigez la désallocation des colonnes à gratter, ne réutilisez pas un enfant comme sortie <code>ConstantVectorExpression</code>
Rétroportage	HIVE-22269 : Corrige le nombre de réducteurs erronés dans les requêtes d'insertion avec partition dynamique en raison de l'absence de statistiques provoquée par -20703. HIVE
Rétroportage	HIVE-22891 : Ignorer <code>PartitionDesc</code> l'extraction <code>CombineHiveRecord</code> en mode non-exécution LLAP
Rétroportage	HIVE-23804 : Ajout d'une base de données par défaut pour les tables spécifiques aux statistiques des colonnes dans le schéma de métastore Hive afin de les rendre rétrocompatibles
Rétroportage	HIVE-25277 : Suppression lente de la partition Hive pour les magasins d'objets Cloud avec des coûts élevés <code>ListFiles</code>

Type	Description
Rétroportage	HIVE-19202 : CBO échec dû à une entrée. NullPointerException HiveAggregate isBucketedInput()
Rétroportage	HIVE-19048 : Fix Beeline : les erreurs Initscript sont ignorées
Rétroportage	HIVE-21085 : Le registre des vues matérialisées démarre une session tez non externe
Rétroportage	HIVE-21675 : CREATE VIEW IF NOT EXISTS renvoie une erreur plutôt que « OK » si la vue existe déjà. Il s'agit d'une régression par rapport à Hive 2.
Rétroportage	HIVE-21646 : Tez : empêche d'échapper au contexte TezTasks de journalisation des threads
Rétroportage	HIVE-22054 : Évitez les listes récursives pour vérifier si un répertoire est vide
Rétroportage	HIVE-16587 : NPE lors de l'insertion de types complexes avec des valeurs nulles imbriquées
Rétroportage	HIVE-22647 : Active le pool de sessions par défaut
Rétroportage	HIVE-13288 : Message d'exception confus dans. DagUtils localizeResource
Rétroportage	HIVE-23870 : Optimisez les conversions de texte multiples dans. WritableHiveCharObjectInspector getPrimitiveJavaObject/HiveCharWritable

Type	Description
Rétroportage	HIVE-21498 : Mise à niveau de Thrift vers la version 0.13.0
Rétroportage	HIVE-24378 : Les espaces de début et de fin ne sont pas supprimés avant la conversion décimale
Rétroportage	HIVE-21341 : Valeurs par défaut raisonnables : hive.server2.idle.operation.timeout et hive.server2.idle.session.timeout sont trop élevés
Rétroportage	HIVE-22465 : Ajout de la configuration SSL dans TezConfigurationFactory
Rétroportage	HIVE-24710 : Optimisation de PTF l'itération pour le nombre (*) afin de réduire le coût des E/S CPU
Rétroportage	HIVE-15406 : Envisagez de vectoriser la nouvelle fonction « trunc »
Rétroportage	HIVE-21541 : Corrige les en-têtes asf manquants dans -15406 HIVE
Rétroportage	HIVE-24808 : Dates analysées dans le cache
Rétroportage	HIVE-24746 : PTF : TimestampValueBoundaryScanner peut être optimisé pendant le calcul de la portée
Rétroportage	HIVE-25059 : L'événement Alter est converti pour être renommé lors de la réplication
Rétroportage	HIVE-25142 : Le remaniement dans la table de hachage rapide de la jointure cartographique entraîne la corruption des clés volumineuses

Type	Description
Rétroportage	HIVE-23756 : Ajout de contraintes supplémentaires au fichier package.jdo
Rétroportage	HIVE-25150 : Les caractères de tabulation ne sont pas supprimés avant la conversion décimale, comme le caractère d'espace qui est corrigé dans le cadre de -24378 HIVE
Rétroportage	HIVE-25093 : date_format () renvoie la sortie uniquement dans UDF le fuseau horaire UTC
Rétroportage	HIVE-25268 : date_format udf renvoie des résultats erronés pour les dates antérieures à 1900 si le fuseau horaire local est autre que UTC
Rétroportage	HIVE-25338 : AIOBE en configuration UDF si l'entrée est vide
Rétroportage	HIVE-22400 : UDF minute avec retour de l'heure NULL
Rétroportage	HIVE-25058 : PTF : TimestampValueBoundaryScanner peut être optimisé lors du calcul de la plage pt2 - isDistanceGreater
Rétroportage	HIVE-25449 : datediff () donne un mauvais résultat lorsqu'il est exécuté dans une tâche tez avec un autre fuseau horaire UTC
Rétroportage	HIVE-23688 : Vectorisation : IndexArrayOutOfBoundsException pour une colonne de type de carte qui inclut une valeur nulle
Rétroportage	HIVE-22247 : HiveHFile OutputFormat lance FileNotFoundException lorsque la sortie de tâche de la partition est vide

Type	Description
Rétroportage	HIVE-25570 : Hive doit envoyer le URL chemin complet pour obtenir l'autorisation de la commande insert overwrite location
Rétroportage	HIVE-22903 : La vectorisation row_number () réinitialise le numéro de ligne après un lot en cas d'expression constante dans la clause de partition
Rétroportage	HIVE-25549 : Mauvais résultats pour la fonction de fenêtre avec une expression dans une clause PARTITION BY ou BY ORDER
Rétroportage	HIVE-25579 : remplace les ajouts plutôt que de LOAD les remplacer
Rétroportage	HIVE-25659 : Les requêtes SQL directes de Metastore avec IN/ (IN) doivent être divisées NOT en fonction du nombre maximum de paramètres autorisés par la base de données SQL
Rétroportage	HIVE-20502 : Correction NPE lors de l'exécution de skewjoin_mapjoin10.q lorsque les statistiques des colonnes sont utilisées.
Rétroportage	HIVE-25765 : la propriété skip.header.line.count ignore les lignes de chaque bloc lorsque la taille du fichier est plus grande FetchOperator
Bogue	NPECorrectif lors de l'insertion dans certains scénarios lorsque hive.stats.column.autogather et hive.groupby.skewindata les deux sont activés

Type	Description
Bogue	Corriger NPE lorsque <code>mapred.tasktracker.expiry.interval</code> la valeur n'est pas définie

Amazon EMR 6.9.0 - Notes de mise à jour de Hive

Amazon EMR 6.9.0 - Changements apportés à Hive

Type	Description
Upgrade	Mise à niveau de Jetty vers 9.4.48.v20220622
Upgrade	Prise en charge pour Hadoop 3.3.3.
Fonctionnalité	Intégration d'Amazon EMR Hive à Lake Formation pour les charges de travail interactives utilisant. GCSC API
Fonctionnalité	Intégration d'Amazon EMR Hive à Iceberg.
Amélioration	Activez SSL en HiveServer 2 lorsque le chiffrement en transit est activé à l'aide des configurations EMR de sécurité Amazon.
Amélioration	Activez le validateur optimisé Hive EMRFS Amazon S3 par défaut. Pour plus d'informations, veuillez consulter Activation du validateur optimisé Hive EMRFS S3 .
Amélioration	Ajoutez HiveHBaseTableInputFormatV2 qui hérite uniquement de la version mappée de InputFormat pour corriger SPARK -34210. Définissez <code>hive.hbase.inputformat.v2</code> sur <code>true</code> pour l'utiliser.

Type	Description
Amélioration	Attendez que TeZam se lance en arrière-plan avec hive.cli.tez.session.async au lieu de l'arrêter et d'en lancer un nouveau immédiatement. Utilisez <code>hive.emr.cli.tez.session.open.timeout</code> pour définir ce délai en secondes.
Amélioration	Ajoutez l'option hive.conf.restricted.list.append pour ajouter les configurations séparées par des virgules à la liste des configurations restreintes existante <code>hive.conf.restricted.list</code> .
Amélioration	Message d'erreur plus clair lorsque la requête Hive échoue car l'emplacement n'est pas défini pour la base de données.
Rétroportage	HIVE-24484 : mise à niveau de Hadoop vers la version 3.3.1 et de Tez vers la version 0.10.2
Rétroportage	HIVE-22398 : Supprime la gestion des YARN files d'attente via. ShimLoader
Rétroportage	HIVE-23190 : LLAP : modifier IndexCache pour transmettre l'objet du système de fichiers à. TezSpillRecord
Rétroportage	HIVE-22185 : HADOOP -15832 posera des problèmes lors des tests utilisant des clusters. MiniYarn
Rétroportage	HIVE-21670 : Remplacement par dépendanc <code>emockito-all</code> . <code>mockito-core</code>
Rétroportage	HIVE-24542 : Préparez la goyave pour les améliorations.

Type	Description
Rétroportage	HIVE-23751 : QTest : Remplacez la <code>#mkdirs()</code> méthode par <code>ProxyFile System</code> pour l'aligner après -16582. HADOOP
Rétroportage	HIVE-21603 : Préparation de Java 11 : mise à jour de la version de PowerMock.
Rétroportage	HIVE-24083 : <code>hcatalog</code> erreur dans Hadoop 3.3.0 : type d'authentification requis.
Rétroportage	HIVE-24282 : L'affichage des colonnes ne doit pas trier les colonnes de sortie sauf mention explicite.
Rétroportage	HIVE-20656 : Valeurs par défaut raisonnables : les configurations de mémoire d'agrégation de cartes sont trop agressives.
Rétroportage	HIVE-25443 : Arrow SerDe ne peut pas sérialiser/désérialiser les types de données complexes lorsqu'il y a plus de 1024 valeurs
Rétroportage	HIVE-19792 : Mettez à jour <code>orc</code> vers la version 1.5.2 et activez les tests d'évolution du schéma <code>decimal_64</code> .
Rétroportage	HIVE-20437 : Gérez l'évolution du schéma à partir de nombres flottants, doubles et décimaux.
Rétroportage	HIVE-21987 : Hive ne parvient pas à lire Parquet <code>int32</code> annoté en décimal.
Rétroportage	HIVE-20038 : Mise à jour des requêtes sur les lancements de tables non partitionnées ou non partitionnées. NPE

Amazon EMR 6.9.0 - Problèmes connus liés à Hive

- Avec Amazon EMR 6.6.0 à 6.9.x, les INSERT requêtes comportant une partition dynamique et une clause ORDER BY ou SORT BY auront toujours deux réducteurs. Ce problème est dû à la OSS modification [HIVE-20703](#), qui soumet l'optimisation des partitions de tri dynamique à une décision basée sur les coûts. Si votre charge de travail ne nécessite pas le tri des partitions dynamiques, nous vous recommandons de définir la propriété `hive.optimize.sort.dynamic.partition.threshold` sur -1 pour désactiver la nouvelle fonctionnalité et obtenir le nombre de réducteurs correctement calculé. Ce problème est résolu dans OSS Hive dans le cadre de la version [HIVE-22269](#) et dans Amazon 6.10.0. EMR

Amazon EMR 6.8.0 - Notes de mise à jour de Hive

Amazon EMR 6.8.0 - Changements apportés à Hive

Type	Description
Amélioration	Réduisez les appels au système de fichiers dans la commande <code>msck</code> . Améliorations des performances (environ 15 à 20 fois sur plus de 10 000 partitions)
Rétroportage	HIVE-20678 : <code>HiveHBaseTableOutputFormat</code> doit implémenter <code>HiveOutputFormat</code> pour garantir la compatibilité
Rétroportage	HIVE-21040 : <code>msck</code> fait une liste inutile de fichiers au dernier niveau de l'arborescence des répertoires
Rétroportage	HIVE-21460 : Le chargement des données suivi d'une requête <code>select *</code> entraîne des résultats incorrects
Rétroportage	HIVE-21660 : Mauvais résultat lors de l'utilisation de tout l'union et de la vue ultérieure avec <code>explode</code>

Type	Description
Rétroportage	HIVE-22505 : ClassCastException causé par une mauvaise sélection d'opérateur vectorisé
Rétroportage	HIVE-22513 : La propagation constante de la colonne coulée lors des opérations de filtrage peut entraîner des résultats incorrects
Rétroportage	HIVE-23435 : Le résultat de la jointure externe complète est l'absence de lignes
Rétroportage	HIVE-24209 : Conversion incorrecte des arguments de recherche pour une NOT BETWEEN opération lorsque la vectorisation est activée
Rétroportage	HIVE-24934 : aucune VectorizedExpressions annotation n'est nécessaire dans G Check enericUDFSQCount
Rétroportage	HIVE-25278 : HiveProjectJoinTransposeRule peut effectuer des transformations non valides avec des expressions de fenêtrage
Rétroportage	HIVE-25505 : Résultats incorrects avec en-tête. skip.header.line.count si la première ligne est vide
Rétroportage	HIVE-26080 : Mise à niveau d'accumulo-core vers la version 1.10.1
Rétroportage	HIVE-26235 : OR La condition sur la colonne binaire renvoie un résultat vide
Bogue	Corriger les journaux d'avertissement SLF4J relatifs aux liaisons multiples dans stderr lors du lancement

Type	Description
Bogue	Corrigez l'échec de la SHOW TABLE EXTENDED requête avec une erreur Wrong FS lorsque la partition et la table se trouvent sur des systèmes de fichiers différents.

Amazon EMR 6.8.0 - Problèmes connus liés à Hive

- Avec Amazon EMR 6.6.0 à 6.9.x, les INSERT requêtes comportant une partition dynamique et une clause ORDER BY ou SORT BY auront toujours deux réducteurs. Ce problème est dû à la OSS modification [HIVE-20703](#), qui soumet l'optimisation des partitions de tri dynamique à une décision basée sur les coûts. Si votre charge de travail ne nécessite pas le tri des partitions dynamiques, nous vous recommandons de définir la propriété `hive.optimize.sort.dynamic.partition.threshold` sur -1 pour désactiver la nouvelle fonctionnalité et obtenir le nombre de réducteurs correctement calculé. Ce problème est résolu dans OSS Hive dans le cadre de la version [HIVE-22269](#) et dans Amazon 6.10.0. EMR

Amazon EMR 6.7.0 - Notes de mise à jour de Hive

Amazon EMR 6.7.0 - Changements apportés à Hive

Type	Description
Fonctionnalité	Intégration d'Amazon EMR Hive avec LakeFormation.
Fonctionnalité	Journalisation d'audit supplémentaire pour le validateur optimisé Hive EMRFS Amazon S3. Configuration Hive : <code>hive.blobstore.output.committer.logging</code> , par défaut : <code>false</code>
Fonctionnalité	Le répertoire cible supprimé lors de l'insertion est remplacé par un résultat de sélection vide dans une table/partition statique non partitionnée afin qu'il se comporte de la même manière

Type	Description
	que Hive 2.x. Configuration Hive : <code>hive.emr.iow.clean.target.dir</code> , par défaut : <code>false</code>
Bogue	Correction d'un échec de requête intermittent lors de l'utilisation du validateur optimisé Hive EMRFS Amazon S3 avec tri des compartiments de partition.
Upgrade	La version de Hive a été mise à niveau vers 3.1.3. Reportez-vous aux notes de mise à jour d'Apache Hive 3.1.3 pour plus de détails.
Upgrade	Parquet mis à niveau vers la version 1.12.2 .
Rétroportage	HIVE-20065 : Metastore ne doit pas s'appuyer sur Jackson 1.x
Rétroportage	HIVE-20071 : Migrer vers Jackson 2.x et empêcher l'utilisation
Rétroportage	HIVE-20607 : TxnHandler doit être utilisé PreparedStatement pour exécuter des requêtes directes SQL
Rétroportage	HIVE-20740 : Supprime le verrouillage global. ObjectStore setConf méthode
Rétroportage	HIVE-20961 : Suppression de la mise en œuvre NVL
Rétroportage	HIVE-22059 : le fichier hive-exec ne contient pas la bibliothèque Jackson (fasterxml)
Rétroportage	HIVE-22351 : Correction d'une utilisation incorrecte des threads ObjectStore dans TestObjectStore

Type	Description
Rétroportage	HIVE-23534 : NPE dans RetryingMetaStoreC lient #invoke lors d'une capture sans message MetaException
Rétroportage	HIVE-24048 : Harmoniser les composants Jackson avec la version 2.10.latest - Hive
Rétroportage	HIVE-24768 : <u>Utilisez</u> jackson-bom partout pour le remplacement de version
Rétroportage	HIVE-24816 : Mise à niveau de Jackson vers la version 2.10.5.1 ou 2.11.0+ en raison de -2020-25649 CVE
Rétroportage	HIVE-25971 : L'arrêt de la tâche Tez est retardé car le pool de threads mis en cache n'est pas fermé
Rétroportage	HIVE-26036 : NPE causé par getMTable () dans ObjectStore

Amazon EMR 6.7.0 - Problèmes connus liés à Hive

- Les requêtes comportant des fonctions de fenêtrage sur la même colonne que la jointure peuvent entraîner des transformations non valides, comme indiqué dans [HIVE-25278](#), et entraîner des résultats incorrects ou des échecs de requête. Une solution de contournement serait de désactiver ces requêtes CBO au niveau de la requête. Le correctif sera disponible dans une EMR version Amazon ultérieure à la version 6.7.0. Pour plus d'informations, contactez AWS le support.
- Avec Amazon EMR 6.6.0 à 6.9.x, les INSERT requêtes comportant une partition dynamique et une clause ORDER BY ou SORT BY auront toujours deux réducteurs. Ce problème est dû à la OSS modification [HIVE-20703](#), qui soumet l'optimisation des partitions de tri dynamique à une décision basée sur les coûts. Si votre charge de travail ne nécessite pas le tri des partitions dynamiques, nous vous recommandons de définir la propriété `hive.optimize.sort.dynamic.partition.threshold` sur `-1` pour désactiver la nouvelle

fonctionnalité et obtenir le nombre de réducteurs correctement calculé. Ce problème est résolu dans OSS Hive dans le cadre de la version [HIVE-22269](#) et dans Amazon 6.10.0. EMR

Amazon EMR 6.6.0 - Notes de mise à jour de Hive

Amazon EMR 6.6.0 - Changements apportés à Hive

Type	Description
Upgrade	Mettez à niveau de Parquet vers la version 1.12.1 .
Upgrade	Mettre à niveau la version Jetty Jars vers 9.4.43.v20210629
Bogue	Correction d'un problème qui provoquait l'installation de Hive sur tous les nœuds de tâche/ cœur alors LLAP qu'il était activé sur un cluster Hive.
Rétroportage	HIVE-25942 : Mise à niveau de commons-io vers la version 2.8.0 en raison de -2021-29425 CVE
Rétroportage	HIVE-25726 : amélioration de la vitesse à 2,3 en raison de -2020-13936 CVE
Rétroportage	HIVE-25680 : Autoriser le HiveMetastore serveur #get_table_meta API à utiliser l'un des modèles d' HiveMetastoreautorisation.
Rétroportage	HIVE-25554 : Mise à niveau de la version de la flèche vers la version 0.15
Rétroportage	HIVE-25242 : La requête s'exécute extrêmement lentement avec vectorized.adaptor = chosen
Rétroportage	HIVE-25085 : MetaStore Les clients ne sont plus partagés entre les sessions.

Type	Description
Rétroportage	HIVE-24827 : La requête d'agrégation Hive renvoie des résultats incorrects pour les fichiers autres que du texte.
Rétroportage	HIVE-24683 : getFileId Hadoop23Shims est sujet à des chemins inexistant NPE
Rétroportage	HIVE-24656 : CBO échoue pour les requêtes dont la valeur est nulle sur les types de carte et de tableau
Rétroportage	HIVE-24556 : Optimisé DefaultGraphWalker pour les cas sans petit-enfant
Rétroportage	HIVE-24408 : Mise à niveau de Parquet vers la version 1.11.1
Rétroportage	HIVE-24391 : Corriger les FIX TestOrcFile défaillances dans la branche 3.1
Rétroportage	HIVE-24362 : le traitement des AST arbres n'est pas optimal pour les arbres comportant un grand nombre de nœuds
Rétroportage	HIVE-24316 : Mise à niveau de la version 1.5.6 à ORC la version 1.5.8 dans la branche 3.1
Rétroportage	HIVE-24307 : Beeline avec le fichier de propriétés et le paramètre -e échoue
Rétroportage	HIVE-24245 : vectorisé PTF avec un nombre et une partition distincte produisant des résultats incorrects.
Rétroportage	HIVE-24224 : Correction du saut d'en-tête/ pied de page pour Hive on Tez sur un fichier compressé

Type	Description
Rétroportage	HIVE-24157 : Le mode strict échoue lors de l'CASThorodatage ↔ numérique
Rétroportage	HIVE-24113 : NPE en G enericUDFTo UnixTimeStamp
Rétroportage	HIVE-23987 : Mise à niveau de la version de la flèche vers la version 0.11.0
Rétroportage	HIVE-23972 : Ajouter un identifiant de client externe au client externe LLAP
Rétroportage	HIVE-23806 : Évitez d'effacer les états des statistiques des colonnes dans toutes les partitions au cas où le schéma serait étendu. Cela améliore l'exécution de l'instruction alter table add columns.
Rétroportage	HIVE-23779 : BasicStatsTask Les informations ne sont pas imprimées dans la console Beeline
Rétroportage	HIVE-23306 : RESET la commande ne fonctionne pas s'il existe une configuration définie par le système. getProperty
Rétroportage	HIVE-23164 : Le serveur n'est pas correctement arrêté à cause de threads autres que les démons
Rétroportage	HIVE-22967 : Support hive.reloadable.aux.jars.path pour Hive on Tez
Rétroportage	HIVE-22934 : Les compteurs de log interactifs du serveur Hive sont envoyés au flux d'erreurs
Rétroportage	HIVE-22901 : La substitution de variables peut conduire à des références non OOM circulaires

Type	Description
Rétroportage	HIVE-22769 : Résultats de requête incorrects et échec de requête lors de la génération fractionnée pour les fichiers texte compressés
Rétroportage	HIVE-22716 : La lecture ByteBuffer est interrompue ParquetFooterInputFromCache
Rétroportage	HIVE-22648 : Mise à niveau de Parquet vers la version 1.11.0
Rétroportage	HIVE-22640 : Decimal64 ColumnVector : ClassCastException lorsque le type de colonne de partition est Decimal
Rétroportage	HIVE-22621 : cas de test instable :. TestLlapSignerImpl testSigning
Rétroportage	HIVE-22533 : Corrige les vulnérabilités possibles de l'interface utilisateur Web des LLAP démons
Rétroportage	HIVE-22532 : PTFPPD peut pousser la limite de manière incorrecte via Rank/ fonction DenseRank
Rétroportage	HIVE-22514 : HiveProtoLoggingHook peut consommer beaucoup de mémoire
Rétroportage	HIVE-22476 : La fonction Hive datediff fournit des résultats incohérents lorsque hive.fetch.h.task.conversion est défini sur none
Rétroportage	HIVE-22429 : Les tables clusterisées migrées à l'aide de bucketing_version 1 sur Hive 3 utilisent bucketing_version 2 pour les insertions

Type	Description
Rétroportage	HIVE-22412 : StatsUtils lancer NPE lorsque vous expliquez
Rétroportage	HIVE-22360 : MultiDelimitSerDe renvoie des résultats erronés dans la dernière colonne lorsque le fichier chargé comporte plus de colonnes que celles du schéma de table
Rétroportage	HIVE-22332 : Hive doit garantir des paramètres d'évolution de schéma valides depuis -540 ORC
Rétroportage	HIVE-22331 : unix_timestamp sans argument renvoie l'horodatage en millisecondes au lieu de secondes
Rétroportage	HIVE-2275 : OperationManager queryIdOperationne nettoie pas correctement plusieurs queryIds
Rétroportage	HIVE-22273 : Le contrôle d'accès échoue lorsqu'un répertoire temporaire est supprimé
Rétroportage	HIVE-22270 : Mise à niveau de commons-io vers la version 2.6
Rétroportage	HIVE-22241 : Implémentation UDF pour interpréter la date/l'horodatage à l'aide de sa représentation interne et du calendrier hybride grégorien-julien
Rétroportage	HIVE-22241 : Implémentation UDF pour interpréter la date/l'horodatage à l'aide de sa représentation interne et de son hybride grégorien-julien

Type	Description
Rétroportage	HIVE-22232 : NPE lorsque hive.order.columna lignment est défini sur false
Rétroportage	HIVE-22231 : Échec de l'écriture d'une requête Hive de grande taille via Knox en raison d'un tube cassé
Rétroportage	HIVE-22221 : Client externe Llap - Nécessité de réduire le nombre LlapBaselInputFormat getSplits
Rétroportage	HIVE-22208 : Le nom de colonne avec le mot clé réservé n'est pas échappé lorsque la requête incluant une jointure sur une table avec une colonne de masque est réécrite
Rétroportage	HIVE-22197 : Common Merge Join lance une exception de conversion de classe.
Rétroportage	HIVE-22170 : from_unixtime et unix_timestamp doivent utiliser le fuseau horaire de la session utilisateur
Rétroportage	HIVE-22169 : Tez : SplitGenerator essaie de rechercher des fichiers de plan qui n'existeront pas pour Tez
Rétroportage	HIVE-22168 : Supprime la journalisation très coûteuse du hotpath du cache LLAP
Rétroportage	HIVE-22161 UDF : FunctionRegistry synchronise sur org.apache.hadoop.hive.q1.udf.UDFTypeclasse

Type	Description
Rétroportage	HIVE-22120 : Correction de résultats erronés ou d'une ArrayOutOfBound exception dans les jointures extérieures gauches de la carte en fonction de conditions limites spécifiques
Rétroportage	HIVE-22115 : Empêche la création d'un appendice de routage de requêtes si la propriété est définie sur false
Rétroportage	HIVE-22113 : Empêcher l'LLAParrêt des activités connexes AMReporter RuntimeException
Rétroportage	HIVE-22106 : Supprime la synchronisation des requêtes croisées pour la partition-eval
Rétroportage	HIVE-22099 : Plusieurs dates liées à des dates ne UDFs peuvent pas gérer correctement les dates juliennes depuis -20007 HIVE
Rétroportage	HIVE-22037 : HS2 devrait être enregistré lors de l'arrêt en raison de OOM
Rétroportage	HIVE-21976 : Le décalage doit être nul au lieu de zéro dans Calcite HiveSortLimit
Rétroportage	HIVE-21924 : Diviser les fichiers texte même s'il existe un en-tête/pied de page
Rétroportage	HIVE-21913 : GenericUDTFGet Splits doit gérer les noms d'utilisateur de la même manière que LLAP
Rétroportage	HIVE-21905 : Amélioration des génériques dans la classe FetchOperator

Type	Description
Rétroportage	HIVE-21902 : HiveServer 2 interfaces utilisateur : l'en-tête de réponse Jetty nécessite X-Frame-Options
Rétroportage	HIVE-21888 : Définissez hive.parquet.timesamp.skip.conversion par défaut sur true
Rétroportage	HIVE-21868 : Vectoriser... CAST FORMAT
Rétroportage	HIVE-21864 : # LlapBaseInputFormat closeAll
Rétroportage	HIVE-21863 : Améliorer le casting de type Vectorizer pour l'expression WHEN
Rétroportage	HIVE-21862 : ORC ppd produit un résultat erroné avec l'horodatage
Rétroportage	HIVE-21846 : Créez un thread dans Tezam qui récupère périodiquement les métriques LlapDaemon
Rétroportage	HIVE-21837 : MapJoin lance une exception lorsque la colonne sélectionnée a des valeurs complètement nulles
Rétroportage	HIVE-21834 : Évitez les appels inutiles pour simplifier les conditions du filtre
Rétroportage	HIVE-21832 : Nouvelles mesures pour obtenir le temps moyen de file d'attente, de service et de réponse
Rétroportage	HIVE-21827 : appels multiples dans la méthode SemanticAnalyzer Do Not Go Through getTableObject ByName

Type	Description
Rétroportage	HIVE-21822 : Exposez les LlapDaemon métriques via une nouvelle méthode API
Rétroportage	HIVE-21818 CBO : La copie génère du trafic de TableRelOptHiveTable métastore
Rétroportage	HIVE-21815 : Les statistiques du ORC fichier sont analysées deux fois
Rétroportage	HIVE-21805 : HiveServer 2 : Utilisez le fast ShutdownHookManager APIs
Rétroportage	HIVE-21799 : activé DynamicPartitionPruningOptimization, lorsque la clé NullPointerException de jointure se trouve sur la colonne d'agrégation
Rétroportage	HIVE-21794 : Ajouter des paramètres de vue matérialisée à sqlStdAuth SafeVarNameRegexes
Rétroportage	HIVE-21768 : JDBC : Supprime le préfixe d'union par défaut pour les requêtes non jointes UNION
Rétroportage	HIVE-21746 : ArrayIndexOutOfBoundsException lors d'une jointure par hachage partitionnée dynamiquement, désactivée CBO
Rétroportage	HIVE-21717 : Impossible de renommer le répertoire dans la tâche de déplacement.
Rétroportage	HIVE-21685 : Mauvaise simplification dans une requête contenant plusieurs clauses IN

Type	Description
Rétroportage	HIVE-21681 : La description formatée affiche des informations incorrectes pour plusieurs clés primaires
Rétroportage	HIVE-21651 : Déplace le serveur protobuf dans hive-exec.
Rétroportage	HIVE-21619 : Imprimer le type d'horodatage sans précision dans Explain Extended SQL
Rétroportage	HIVE-21592 : OptimizedSql n'apparaît pas lorsque l'expression contient CONCAT
Rétroportage	HIVE-21576 : Présentez... CAST FORMATet liste limitée des formats de date/heure 2016 SQL
Rétroportage	HIVE-21573 : Le transport binaire doit ignorer le principal si l'authentification est définie sur delegationToken
Rétroportage	HIVE-21550 : les TestObjectStore tests sont flous - Impossible d'obtenir un verrou dans le délai demandé
Rétroportage	HIVE-21544 : La propagation constante corrompt les expressions coalesce/case/when lors du pliage
Rétroportage	HIVE-21539 : GroupBy + où la clause sur la même colonne entraîne une réécriture incorrecte de la requête
Rétroportage	HIVE-21538 : Beeline : source du mot de passe même si le lecteur de console n'est pas passé au paramètre de connexion

Type	Description
Rétroportage	HIVE-21509 : LLAP peut mettre en cache des vecteurs de colonne corrompus et renvoyer un résultat de requête erroné
Rétroportage	HIVE-21499 : ne doit pas supprimer la fonction du registre si la commande de création a échoué avec AlreadyExistsException
Rétroportage	HIVE-21496 : Le dimensionnement automatique de la mémoire tampon non ordonnée peut déborder
Rétroportage	HIVE-21468 : distinction majuscules/minuscules dans les noms d'identifiant pour JDBC le gestionnaire de stockage
Rétroportage	HIVE-21467 : Supprimer les importations JUnit.Framework.Assert obsolètes
Rétroportage	HIVE-21435 : LlapBaselInputFormat devrait obtenir le numéro de tâche dans la configuration TASK _ ATTEMPT _ ID si elle est présente, lors de la construction SubmitWorkRequestP roto
Rétroportage	HIVE-21389 : La distribution Hive manque javax.ws.rs-api.jar après -21247 HIVE
Rétroportage	HIVE-21385 : Autoriser la désactivation du transfert des calculs non séparables vers les sources JDBC
Rétroportage	HIVE-21383 : gestionnaire JDBC de stockage : utilisez le catalogue et le schéma pour récupérer les tables si cela est spécifié

Type	Description
Rétroportage	HIVE-21382 : Optimisation de la réduction des groupes par clés - les clés ne sont pas réduites dans la requête 23
Rétroportage	HIVE-21362 : Ajoutez un format d'entrée et serde pour lire à partir des fichiers protobuf.
Rétroportage	HIVE-21340 CBO : Taillez les colonnes non clés alimentant un SemiJoin
Rétroportage	HIVE-21332 : Purger les tampons non verrouillés au lieu des tampons verrouillés
Rétroportage	HIVE-21329 : Taille de la mémoire tampon de sortie non ordonnée du runtime Tez personnalisé en fonction du pipeline de l'opérateur
Rétroportage	HIVE-21295 : StorageHandler doit convertir la date en chaîne en utilisant la convention Hive
Rétroportage	HIVE-21294 : Vectorisation : Shuffle à 1 réducteur peut ignorer les fonctions de hachage des objets
Rétroportage	HIVE-21255 : Supprimer QueryConditionBuilder JdbcStorageHandler
Rétroportage	HIVE-21253 : Support dans DB2 JDBC StorageHandler
Rétroportage	HIVE-21232 : LLAP : Ajout d'un fournisseur d'affinité partagée adapté aux cache-miss
Rétroportage	HIVE-21214 : MoveTask : À utiliser à la attemptId place de la taille du fichier pour la déduplication des fichiers compareTempOr DuplicateFiles

Type	Description
Rétroportage	HIVE-21184 : Ajouter, expliquer et expliquer un CBO plan formaté avec des informations sur les coûts
Rétroportage	HIVE-21182 : Ignorer la configuration de Hive Scratch Dir lors de la planification
Rétroportage	HIVE-21171 : Ignorez la création de scratch pour Tea If Activé RPC
Rétroportage	HIVE-21126 : Autoriser les requêtes au niveau de la session dans # LlapBaseInputFormat getSplit
Rétroportage	HIVE-21107 : Erreur « Impossible de trouver le champ » lors d'une jointure par hachage partitionnée dynamiquement
Rétroportage	HIVE-21061 : la CTAS requête échoue avec une source IllegalStateException vide
Rétroportage	HIVE-21041 :NPE, ParseException en obtenant le schéma à partir du plan logique
Rétroportage	HIVE-21013 : Impossible de JdbcStorageHandler trouver la colonne de partition dans Oracle
Rétroportage	HIVE-21006 : Étendre SharedWorkOptimizer pour supprimer les semi-jointures lorsqu'il existe une opportunité de réutilisation
Rétroportage	HIVE-20992 : Divisez le fichier de configuration hive.metastore.dbaccess.ssl.properties en configurations plus significatives

Type	Description
Rétroportage	HIVE-20989 : JDBC - Le journal GetOperationStatus + peut bloquer la progression de la requête via le mode veille
Rétroportage	HIVE-20988 : Résultats erronés pour les requêtes de groupe par avec une clé primaire sur plusieurs colonnes
Rétroportage	HIVE-20985 : Si les entrées de l'opérateur sélectionné sont des colonnes temporaires, la vectorisation peut réutiliser certaines d'entre elles en sortie
Rétroportage	HIVE-20978 : « hive.jdbc.* » doit être ajouté à sqlStdAuth SafeVarNameRegexes
Rétroportage	HIVE-20953 : Supprime une fonction du registre des fonctions lorsqu'elle ne peut pas être ajoutée au métastore lors de sa création.
Rétroportage	HIVE-20952 : Nettoyage du fichier .java VectorizationContext
Rétroportage	HIVE-20951 : LLAP : Réglez toujours Xms à 50 %
Rétroportage	HIVE-20949 : Améliorer l'estimation de la PKFK cardinalité dans la planification physique
Rétroportage	HIVE-20944 : Les statistiques ne sont pas validées lors de la compilation des requêtes
Rétroportage	HIVE-20940 : Cas de pont dans lesquels la résolution de type de Calcite est plus stricte que celle de Hive.

Type	Description
Rétroportage	HIVE-20937 : La requête JDBC de Postgres échoue avec le message « ne doit pas être négatif » LIMIT
Rétroportage	HIVE-20926 : L'indicateur de réduction des semi-jointures échoue lorsque le nombre d'entrées du filtre Bloom est élevé ou lorsqu'il n'y a aucune statistique
Rétroportage	HIVE-20920 : Utiliser des SQL contraintes pour améliorer l'algorithme de réorganisation des jointures
Rétroportage	HIVE-20918 : Indicateur pour activer/désactiver le transfert du calcul depuis Calcite vers la connexion JDBC
Rétroportage	HIVE-20915 : Rendre l'optimisation des partitions de tri dynamique disponible pour HoS et MR
Rétroportage	HIVE-20910 : L'insertion dans une table à compartiments échoue en raison de l'optimisation dynamique du tri des partitions
Rétroportage	HIVE-20899 : Keytab URI for LLAP YARN Service est limité au support uniquement HDFS
Rétroportage	HIVE-20898 : Pour les fonctions liées au temps, les arguments ne peuvent pas être convertis en un type non nullable
Rétroportage	HIVE-20881 : La propagation constante simplifiée à outrance les projections
Rétroportage	HIVE-20880 : Mettre à jour la valeur par défaut pour <code>hive.stats.filter.in.min.ratio</code>

Type	Description
Rétroportage	HIVE-20873 : Utilisez le hachage Murmur pour VectorHashKeyWrapperTwoLong réduire les collisions de hachage
Rétroportage	HIVE-20868 : La SMB jointure échoue par intermittence lorsqu'un enfant se TezDummyOperator connecte getFinalOp MapRecord Processor
Rétroportage	HIVE-20853 : Exposer . ShuffleHandler registerDagdans le démon Llap API
Rétroportage	HIVE-20850 : Transférez les projections aux tableaux de dimensions si possible
Rétroportage	HIVE-20842 : correction de la logique introduite dans HIVE -20660 pour estimer les statistiques pour le groupe par
Rétroportage	HIVE-20839 : Erreur « Impossible de trouver le champ » lors de la jointure par hachage partitionnée dynamiquement
Rétroportage	HIVE-20835 : L'interaction entre les contraintes et la réécriture de MV peut créer une boucle dans le planificateur Calcite
Rétroportage	HIVE-20834 : QueryResultCache entrées Hive conservant la référence à SemanticAnalyzer la requête mise en cache
Rétroportage	HIVE-20830 : échec de l'assertion JdbcStorageHandler de la requête de plage dans certains cas
Rétroportage	HIVE-20829 : lancers à JdbcStorageHandler portée partagée NPE

Type	Description
Rétroportage	HIVE-20827 : Résultats incohérents pour les tableaux vides
Rétroportage	HIVE-20826 : amélioration de la HiveSemiJoin règle pour convertir la jointure et le groupe situés sur le côté gauche en semi-jointure gauche
Rétroportage	HIVE-20821 : Réécrivez SUM 0 en combinaison + SUM COALESCE
Rétroportage	HIVE-20815 : JdbcRecordReader .next ne doit pas manger d'exception
Rétroportage	HIVE-20813 : udf to_epoch_milli doit également prendre en charge l'horodatage sans fuseau horaire.
Rétroportage	HIVE-20804 : Améliorations supplémentaires apportées au regroupement par optimisation avec contraintes
Rétroportage	HIVE-20792 : L'insertion d'un horodatage avec des zones tronque les données
Rétroportage	HIVE-20788 : La réduction SJ étendue peut revenir en arrière de manière incorrecte lors de la création de filtres
Rétroportage	HIVE-20778 : La réorganisation des jointures ne peut pas être déclenchée si toutes les jointures du plan sont créées par une logique de décorrélation
Rétroportage	HIVE-20772 : enregistre les compteurs par tâche dans CPU LLAP

Type	Description
Rétroportage	HIVE-20768 : Ajout d'une fenêtre Tumbling UDF
Rétroportage	HIVE-20767 : Plusieurs projets entre opérateurs de jointure peuvent affecter la réorganisation des jointures en utilisant des contraintes
Rétroportage	HIVE-20762 : NOTIFICATION _ l'intervalle de LOG nettoyage est codé en dur comme 60 secondes et est trop petit
Rétroportage	HIVE-20761 : La sélection pour la mise à jour de la table notification_sequence présente un intervalle de tentatives et un nombre de tentatives trop faible
Rétroportage	HIVE-20751 : Mise à niveau de la version de la flèche vers la version 0.10.0
Rétroportage	HIVE-20746 : HiveProtoHookLogger ne ferme pas le fichier en fin de journée.
Rétroportage	HIVE-20744 : Utiliser des SQL contraintes pour améliorer l'algorithme de réorganisation des jointures
Rétroportage	HIVE-20740 : Supprime le verrouillage global. ObjectStore setConf méthode. Ce cherrypick rétroporte HIVE -20740 destiné à Hive 3.2 et 4.x vers 3.1.x
Rétroportage	HIVE-20734 : Beeline : lorsque beeline-site.xml est activé et que Hive CLI redirige vers Beeline, il doit utiliser le nom d'utilisateur/mot de passe factice du système au lieu de demander un

Type	Description
Rétroportage	HIVE-20731 : le fichier keystore doit être autorisé JdbcStorageHandler
Rétroportage	HIVE-20720 : Ajouter une option de colonne de partition au gestionnaire JDBC
Rétroportage	HIVE-20719 : L'INSERT instruction échoue après avoir activé l'optimisation et la vectorisation de UPDATE hive.optimize.sort.dynamic.partition
Rétroportage	HIVE-20718 : Ajout d'un pilote Perf CLI avec des contraintes
Rétroportage	HIVE-20716 : Définissez la valeur par défaut de hive.cbo.stats.correlated.multi.key.joins sur true
Rétroportage	HIVE-20712 : HivePointLookupOptimizer devrait extraire les cas profonds
Rétroportage	HIVE-20710 : Le pliage constant ne peut pas créer de constantes nulles sans types
Rétroportage	HIVE-20706 : external_jdbc_table2.q échoue par intermittence
Rétroportage	HIVE-20704 : extension HivePreFilteringRule pour prendre en charge d'autres fonctions
Rétroportage	HIVE-20703 : Placer l'optimisation des partitions de tri dynamique dans le cadre d'une décision basée sur les coûts
Rétroportage	HIVE-20702 : Tenez compte des frais généraux liés aux estimations tenant compte de la structure des données lors de la sélection de mapjoin

Type	Description
Rétroportage	HIVE-20692 : Active le pliage des expressions NOT x IS (NOT) [[]] TRUE FALSE
Rétroportage	HIVE-20691 : Corrigez org.apache.hadoop.hive.cli. TestMiniLlapCliDriver. testCliDriver[cttl]
Rétroportage	HIVE-20682 : L'exécution d'une requête asynchrone peut potentiellement échouer si le partage sessionHive est fermé par le thread principal
Rétroportage	HIVE-20676 : HiveServer 2 : le statut de démon n' PrivilegeSynchronizerest pas défini
Rétroportage	HIVE-20660 : L'estimation des groupes par statistiques pourrait être améliorée en limitant le nombre total de lignes à la table source
Rétroportage	HIVE-20652 : jointure JdbcStorageHandler push de deux sources de données différentes vers le pilote jdbc
Rétroportage	HIVE-20651 : le mot de JdbcStorageHandler passe doit être crypté
Rétroportage	HIVE-20649 : gestionnaire de mémoire compatible LLAP pour les écrivains Orc
Rétroportage	HIVE-20648 : LLAP : Le groupe de vecteurs par opérateur doit utiliser de la mémoire par exécuteur
Rétroportage	HIVE-20646 : La condition du filtre de partition n'est pas transmise à la requête Metastore si elle possède un IS NOT NULL

Type	Description
Rétroportage	HIVE-20644 : Évitez d'exposer des informations sensibles par le biais d'une exception Hive Runtime
Rétroportage	HIVE-20636 : Amélioration de l'estimation du nombre de valeurs nulles après une jointure externe
Rétroportage	HIVE-20632 : La requête avec get_splits UDF échoue si une vue matérialisée est créée sur la table interrogée
Rétroportage	HIVE-20627 : Les requêtes asynchrones simultanées échouent par intermittence et provoquent une fuite de mémoire LockException
Rétroportage	HIVE-20623 : Travail partagé : étendre le partage des entrées du cache Map-Join dans LLAP
Rétroportage	HIVE-20619 : Inclus MultiDelimitSerDe dans 2 par défaut HiveServer
Rétroportage	HIVE-20618 : Lors de la jointure, la sélection BucketMapJoin peut être choisie pour les tables non groupées
Rétroportage	HIVE-20617 : Corrige le type des constantes dans les expressions IN pour qu'elles aient le type correct
Rétroportage	HIVE-20612 : Création d'un nouvel indicateur de corrélation multiclé de jointure pour CBO

Type	Description
Rétroportage	HIVE-20603 : Erreur « Wrong FS » lors de l'insertion dans la partition après avoir modifié l'emplacement de la table dans le système de fichiers
Rétroportage	HIVE-20601 : EnvironmentContext nul dans ALTER _ PARTITION événement dans DbNotificationListener
Rétroportage	HIVE-20583 : Utiliser le nom d'hôte canonique uniquement pour l'authentification Kerberos dans HiveConnection
Rétroportage	HIVE-20582 : Rendre la journalisation des proto hflush dans Hive configurable
Rétroportage	HIVE-20563 : Vectorisation : l'CASEWHE Nexpression échoue lorsque le ELSE typeTHEN/et le type de résultat sont différents
Rétroportage	HIVE-20558 : Remplacer la valeur par défaut de hive.hashtable.key.count.adjustment à 0,99
Rétroportage	HIVE-20552 : Récupère le schéma plus rapidement LogicalPlan
Rétroportage	HIVE-20550 : Switch W ebHCat pour utiliser beeline pour soumettre des requêtes Hive
Rétroportage	HIVE-20537 : Plusieurs colonnes joignent les estimations avec des colonnes non corrélées différentes dans et Hive CBO
Rétroportage	HIVE-20524 : La vérification de l'évolution du schéma est interrompue lors du passage de la version 2 à la version 3 de Hive pour ALTER TABLE VARCHAR DECIMAL

Type	Description
Rétroportage	HIVE-20522 : HiveFilterSetOpTransposeRule peut générer une erreur d'assertion en raison de la nullité des champs
Rétroportage	HIVE-20521 : HS2 doAs =true a un problème d'autorisation avec hadoop.tmp.dir, avec les systèmes de fichiers MR et S3A
Rétroportage	HIVE-20515 : Résultats de requête vides lors de l'utilisation du cache des résultats et du répertoire temporaire des requêtes, répertoire et du cache des résultats dans différents systèmes de fichiers
Rétroportage	HIVE-20508 : Hive ne prend pas en charge les noms d'utilisateur de type « user @realm »
Rétroportage	HIVE-20507 : Beeline : ajoute une commande utilitaire pour récupérer tous les URI du fichier beeline-site.xml
Rétroportage	HIVE-20505 : mise à niveau de org.openjdk.jmh:jmh-core vers la version 1.21
Rétroportage	HIVE-20503 : Utiliser des estimations tenant compte de la structure des données lors de la sélection de mapjoin
Rétroportage	HIVE-20498 : Type de date de support pour la collecte automatique des statistiques des colonnes
Rétroportage	HIVE-20496 : Vectorisation : vectorisée PTF IllegalStateException
Rétroportage	HIVE-20494 : G enericUDFRestrict InformationSchema est cassé après -19440 HIVE

Type	Description
Rétroportage	HIVE-20477 : n' OptimizedSql est pas affiché si l'expression contient INs
Rétroportage	HIVE-20467 : Autoriser IF /IF dans la NOT EXISTS création/suppression du plan EXISTS de ressources
Rétroportage	HIVE-20462 : « CREATE VIEW IF NOT EXISTS » échoue si la vue existe déjà
Rétroportage	HIVE-20455 : Log généré par security.authorization.PrivilegeSynchronizer.exécuter
Rétroportage	HIVE-20439 : Utiliser la limite de mémoire gonflée lors de la sélection des jointures pour llap
Rétroportage	HIVE-20433 : La conversion implicite d'une chaîne en horodatage est lente
Rétroportage	HIVE-20432 : Réécriture dans IN pour BETWEEN les types entiers pour l'estimation des statistiques
Rétroportage	HIVE-20423 : défini NULLS LAST comme ordre nul par défaut
Rétroportage	HIVE-20418 : LLAP IO peut ne pas gérer correctement ORC les fichiers dont l'index de ligne est désactivé pour les requêtes sans colonne sélectionnée
Rétroportage	HIVE-20412 : entrée NPE HiveMetaHook
Rétroportage	HIVE-20406 : La coalescence imbriquée donne des résultats incorrects

Type	Description
Rétroportage	HIVE-20399 : CTAS avec un emplacement de table personnalisé qui n'est pas entièrement qualifié, échoue pour les tables MM
Rétroportage	HIVE-20393 : Réduction des semi-jointures : car se comporte de manière incohérente markSemiJoin DPP
Rétroportage	HIVE-20391 : HiveAggregateReduceFunction sRule peut déduire un type de retour erroné lors de la décomposition de la fonction d'agrégation
Rétroportage	HIVE-20383 : Nom de file d'attente non valide et problèmes de synchronisation dans Hive Proto Events Hook.
Rétroportage	HIVE-20367 : Vectorisation : Support du streaming PTF AVG pour,,, MAX MIN SUM
Rétroportage	HIVE-20366 : Les estimations de statistiques de TPC -DS query78 sont désactivées pour son filtre nul
Rétroportage	HIVE-20364 : Mise à jour par défaut pour hive.map.aggr.hash.min.reduction
Rétroportage	HIVE-20352 : Vectorisation : fonction de regroupement de supports
Rétroportage	HIVE-20347 : hive.optimize.sort.dynamic.partition devrait fonctionner avec partitioned et MV CTAS
Rétroportage	HIVE-20345 : La base de données Drop peut se bloquer si les tables sont supprimées lors d'un autre appel

Type	Description
Rétroportage	HIVE-20343 : Hive 3 : CTAS ne respecte pas les propriétés transactionnelles
Rétroportage	HIVE-20340 : Le druide a besoin d'une information explicite CASTs entre l'horodatage et le moment STRING où la sortie de la fonction d'horodatage est utilisée sous forme de chaîne
Rétroportage	HIVE-20339 : Vectorisation : supprimez une restriction inutile, empêchant ainsi la vectorisation de certaines PTF restrictions RANK
Rétroportage	HIVE-20337 : CachedStore : getPartitionsBy Expr ne remplit pas correctement la liste des partitions
Rétroportage	HIVE-20336 : Politiques de masquage et de filtrage pour les vues matérialisées
Rétroportage	HIVE-20326 : Créez des contraintes avec RELY comme valeur par défaut au lieu de NO RELY
Rétroportage	HIVE-20321 : Vectorisation : réduisez la taille de la mémoire de 1 col à <1 VectorHas hKeyWrapper CacheLine
Rétroportage	HIVE-20320 : Activez l'indicateur hive.optimize.remove.sql_count_check
Rétroportage	HIVE-20315 : Vectorisation : corrigez davantage de problèmes liés aux résultats erronés et évitez les NULL casts/conversions inutiles
Rétroportage	HIVE-20314 : Inclure l'élagage des partitions dans la réécriture des vues matérialisées

Type	Description
Rétroportage	HIVE-20312 : Autoriser les clients Arrow à utiliser les leurs avec BufferAllocator LlapOutputFormatService
Rétroportage	HIVE-20302 : LLAP : l'exécution non vectorisé e dans IO ignore les colonnes virtuelles, y compris __ID ROW
Rétroportage	HIVE-20300 : VectorFileSinkArrowOperator
Rétroportage	HIVE-20299 : course potentielle lors du test unitaire des LLAP signataires
Rétroportage	HIVE-20296 : Améliorez HivePointLookupOptimizerRule pour pouvoir extraire des données de contextes plus sophistiqués
Rétroportage	HIVE-20294 : Vectorisation : correction des problèmes de résultats NULL erronés dans/ COALESCELT
Rétroportage	HIVE-20292 : Mauvais ordre des jointures dans la requête tpcds 93 avec une contrainte principale définie
Rétroportage	HIVE-20290 : Initialisation lente pour ne ArrowColumnarBatchSerDe pas allouer de tampons pendant GetSplits
Rétroportage	HIVE-20281 : SharedWorkOptimizer échec avec « le contenu du cache de l'opérateur et le plan réel diffèrent »
Rétroportage	HIVE-20277 : Vectorisation : les expressions de cas renvoyées ne sont pas prises en charge pour BOOLEAN FILTER

Type	Description
Rétroportage	HIVE-20267 : Extension de WebUI pour inclure un formulaire afin de configurer dynamiquement les niveaux de journalisation
Rétroportage	HIVE-20263 : faute de frappe dans la variable HiveReduceExpressionsWithStatsRule
Rétroportage	HIVE-20260 : NDV d'une colonne ne doit pas être redimensionné lorsque le nombre de lignes est modifié par un filtre sur une autre colonne
Rétroportage	HIVE-20252 : Réduction des semi-jointures : les cycles dus à une branche de semi-jointure peuvent passer inaperçus si une jointure cartographique se trouve sur le côté de la petite table en amont.
Rétroportage	HIVE-20245 : Vectorisation : correction des problèmes de résultats NULL erronés dans/IN BETWEEN
Rétroportage	HIVE-20241 : Support des spécifications de partitionnement dans les déclarations CTAS
Rétroportage	HIVE-20240 : Réduction des semi-jointures : utilisez une variable locale pour vérifier l'état de la table externe
Rétroportage	HIVE-20226 : HMS getNextNotification lancera une exception lorsque la demande maxEvents dépasse le max_rows de la table
Rétroportage	HIVE-20225 : SerDe pour prendre en charge le format binaire Teradata
Rétroportage	HIVE-20213 : Mise à niveau de Calcite vers la version 1.17.0

Type	Description
Rétroportage	HIVE-20212 : Hiveserver2 en mode http émet la métrique Default.general.open_connections de manière incorrecte
Rétroportage	HIVE-20210 : L'optimiseur Simple Fetch devrait permettre de filtrer sur une colonne non MapReduce partitionnée et que la conversion est minimale
Rétroportage	HIVE-20209 : La connexion au métastore échoue lors de la première tentative de vidage de repl
Rétroportage	HIVE-20207 : Vectorisation : correction des problèmes de résultats NULL erronés dans Filtre/Comparaison
Rétroportage	HIVE-20204 : Conversion de type pendant IN
Rétroportage	HIVE-20203 : Arrow SerDe fuit un DirectByteBuffer
Rétroportage	HIVE-20197 : Vectorisation : ajoutez des tests DECIMAL_64, ajoutez l'arithmétique Date/Interval/Horodatage et ajoutez-en d'autres par agrégation GROUP
Rétroportage	HIVE-20193 : n' ccolInfo est pas présent dans le fichier json du plan d'explication
Rétroportage	HIVE-20192 : HS2 avec le métastore intégré, des objets fuient JDOPersistenceManager
Rétroportage	HIVE-20183 : L'insertion depuis une table à compartiments peut entraîner une perte de données si la table source contient un compartiment vide

Type	Description
Rétroportage	HIVE-20177 : Vectorisation : réduction KeyWrapper de l'allocation en mode Streaming GroupBy
Rétroportage	HIVE-20174 : Vectorisation : correction des problèmes de résultats erronés dans NULL les fonctions d'agrégation BY GROUP
Rétroportage	HIVE-20172 : StatsUpdater échec avec une GSS exception lors de la tentative de connexion à une métastore distante
Rétroportage	HIVE-20153 : Count et Sum UDF consomment plus de mémoire dans Hive 2+
Rétroportage	HIVE-20152 : réinitialise l'état de la base de données, lorsque le vidage de repl échoue, afin de renommer la table
Rétroportage	HIVE-20149 : TestHiveCli échec/expiration du délai
Rétroportage	HIVE-20130 : Meilleure journalisation pour le synchroniseur de schémas d'informations
Rétroportage	HIVE-20129 : Revenir à l'évolution du schéma basée sur la position pour les tables orc
Rétroportage	HIVE-2018 : SessionStateUserAuthenticator getGroupNames
Rétroportage	HIVE-20116 : utilise l' TezTask enregistreur parent
Rétroportage	HIVE-20115 : Les tableaux acides ne doivent pas utiliser le scan du pied de page pour l'analyse

Type	Description
Rétroportage	HIVE-20103 : WM : Agrégez uniquement DAG les compteurs si au moins un est utilisé
Rétroportage	HIVE-20101 : B loomKFilter : Évitez d'utiliser entièrement les tableaux d'octets locaux []
Rétroportage	HIVE-20100 OpTraits : Select Optraits devrait cesser lorsqu'une incompatibilité est détectée
Rétroportage	HIVE-20098 : Statistiques : NPE lors de l'obtention des statistiques de partition de la colonne Date
Rétroportage	HIVE-20095 : Correction d'une fonctionnalité permettant de transférer le calcul vers des tables externes JDBC
Rétroportage	HIVE-20093 : LlapOutputFomatService : Utilisation ArrowBuf avec Netty pour la comptabilité
Rétroportage	HIVE-20090 : Étendre la création de filtres de réduction des semi-jointures pour pouvoir découvrir de nouvelles opportunités
Rétroportage	HIVE-20088 : Le chemin de localisation de la configuration Beeline est mal assemblé
Rétroportage	HIVE-20082 : la conversion HiveDecimal en chaîne ne formate pas correctement le décimal
Rétroportage	HIVE-20069 : Correction de la réoptimisation en cas d'optimisation et de semi-jointure DPP
Rétroportage	HIVE-20051 : Ignorer l'autorisation pour les tables temporaires

Type	Description
Rétroportage	HIVE-20044 : Arrow Serde doit enregistrer les valeurs des caractères et gérer correctement les chaînes vides
Rétroportage	HIVE-20028 : La configuration du cache du client Metastore n'est pas utilisée correctement
Rétroportage	HIVE-20025 : Nettoyage des fichiers d'événements créés par HiveProtoLoggingHook
Rétroportage	HIVE-20020 : Le fichier jar de contributions Hive ne doit pas être dans lib
Rétroportage	HIVE-20013 : Ajouter un type de conversion implicite à la date pour la fonction to_date
Rétroportage	HIVE-20011 : Éloignez-vous du mode ajout dans Proto Logging Hook
Rétroportage	HIVE-20005 : acid_table_stats, acid_no_buckets, etc. - modification du résultat de la requête sur la branche
Rétroportage	HIVE-20004 : Une mauvaise échelle utilisée par ConvertDecimal 64 ToDecimal entraîne des résultats incorrects
Rétroportage	HIVE-19995 : Trafic de lignes agrégé pour les tables acides
Rétroportage	HIVE-19993 : Il n'est pas possible d'utiliser un alias de table qui apparaît également sous forme de nom de colonne

Type	Description
Rétroportage	HIVE-19992 : Vectorisation : suivi de HIVE -19951 --> ajout d'un appel à. SchemaEvolution isOnlyImplicitConversion pour désactiver les LLAP E/S codées ORC uniquement lorsque la conversion de type de données n'est pas implicite
Rétroportage	HIVE-19989 : Metastore utilise un nom d'application incorrect pour les métriques HADOOP2
Rétroportage	HIVE-19981 : Les tables gérées converties en tables externes par l' HiveStrictManagedM igration utilitaire doivent être configurées pour supprimer les données lorsque la table est supprimée
Rétroportage	HIVE-19967 : SMB Join : Need Optraits for ala Op PTFOperator GBY
Rétroportage	HIVE-19935 : Session Hive WM interrompue : échec de la mise à jour du nombre de tâches LLAP
Rétroportage	HIVE-19924 : Taguer les tâches distcp exécutées par Repl Load
Rétroportage	HIVE-19891 : l'insertion dans des tables externes avec des répertoires de partitions personnalisés peut entraîner une perte de données
Rétroportage	HIVE-19850 : L'élagage dynamique des partitions dans Tez entraîne l'erreur « Aucune tâche trouvée pour la numérisation de tables »

Type	Description
Rétroportage	HIVE-19806 : Triez les résultats des tests pour éviter les défauts dans les résultats des tests
Rétroportage	HIVE-19770 : Support CBO pour les requêtes comportant plusieurs colonnes identiques dans select
Rétroportage	HIVE-19769 : Création d'objets dédiés pour les noms de bases de données et de tables
Rétroportage	HIVE-19765 : Ajouter des tests spécifiques au parquet à BlobstoreCliDriver
Rétroportage	HIVE-19759 : Test floconneux : # TestRpc testServerPort
Rétroportage	HIVE-19711 : Outil de schéma Refactor Hive
Rétroportage	HIVE-19701 : getDelegationToken FromMetaStore n'a pas besoin d'être synchronisé
Rétroportage	HIVE-19694 : L'instruction Create Materialized View doit vérifier les conflits de noms de MV avant d'exécuter l'instruction MV. SQL
Rétroportage	HIVE-19674 : Regrouper par constantes décimales jusqu'à la table des druides
Rétroportage	HIVE-19668 : Plus de 30 % du tas est gaspillé à cause de la duplication de org.antlr.runtime.CommonTokenet chaînes dupliquées
Rétroportage	HIVE-19663 : génération de rapports d'E/S refactorisés LLAP
Rétroportage	HIVE-19661 : changez Hive UDFs pour utiliser le moteur Re2J regex

Type	Description
Rétroportage	HIVE-19628 : possible dans NPE LLAP testSigning
Rétroportage	HIVE-19568 : HS2 HA actif/passif : interdire la connexion directe à une instance passive HS2
Rétroportage	HIVE-19564 : Vectorisation : correction des problèmes de résultats erronés en NULL arithmétique
Rétroportage	HIVE-19552 : Activez #druidkafkamini_basic .q TestMiniDruidKafkaCliDriver
Rétroportage	HIVE-19432 : GetTablesOperation est trop lent si la ruche contient trop de bases de données et de tables
Rétroportage	HIVE-19360 : CBO : Ajouter un « optimisé SQL » à l'objet QueryPlan
Rétroportage	HIVE-19326 : collecte automatique des statistiques : agrégation incorrecte lors des requêtes UNION
Rétroportage	HIVE-19313 : les TestJdbcWith DBTokenStoreNoDoAs tests échouent
Rétroportage	HIVE-19285 : Ajoutez des journaux aux sous-classes de MetadataOperation
Rétroportage	HIVE-19235 : Mise à jour des fichiers dorés pour les tests Minimr
Rétroportage	HIVE-19104 : Lorsque le test MetaStore est lancé avec une nouvelle tentative, les instances doivent être indépendantes

Type	Description
Rétroportage	HIVE-18986 : Le changement de nom de table exécutera java.lang. StackOverflowError dans dataNucleus si le tableau contient un grand nombre de colonnes
Rétroportage	HIVE-18920 : CBO : Initialise les fournisseurs Janino avant la première requête
Rétroportage	HIVE-18873 : Le fait d'ignorer silencieusement le renvoi du prédicat pour la résonance magnétique HiveInputFormat peut entraîner la production d'un résultat erroné par les gestionnaires de stockage
Rétroportage	HIVE-18871 : erreur d'exécution de hive on tez due à la définition de hive.aux.jars.path sur hdfs ://
Rétroportage	HIVE-18725 : Améliore la gestion des erreurs pour les sous-requêtes en cas de référence de colonne incorrecte
Rétroportage	HIVE-18696 : Les dossiers de partition risquent de ne pas être nettoyés correctement dans la méthode HiveMetaStore .add_partitions_core si un
Rétroportage	HIVE-18453 : ACID : Ajout de la syntaxe « CREATE TRANSACTIONAL TABLE » pour unifier et prendre en charge le parquet ACID ORC
Rétroportage	HIVE-18201 : Désactiver XPROD _ EDGE pour sq_count_chec

Type	Description
Rétroportage	HIVE-18140 : Les statistiques des tables partitionnées peuvent mal tourner dans les statistiques de base
Rétroportage	HIVE-17921 : L'agrégation avec struct in produit un résultat erroné LLAP
Rétroportage	HIVE-17896 : T opNKey : Création d'un opérateur T vectorisable autonome opNKey
Rétroportage	HIVE-17840 : HiveMetaStore mange une exception si. transactionalListeners notifyEvent échouer
Rétroportage	HIVE-17043 : Supprime les colonnes non uniques du groupe par clés si elles ne sont pas référencées ultérieurement
Rétroportage	HIVE-17040 : Élimination conjointe en présence d'une relation FK
Rétroportage	HIVE-16839 : appels déséquilibrés vers openTransaction/commitTransaction lorsque vous modifiez simultanément la même partition
Rétroportage	HIVE-16100 : L'optimiseur de partitions triées dynamiques perd ses opérateurs frères
Rétroportage	HIVE-15956 : StackOverflowError lorsque vous déposez un grand nombre de partitions
Rétroportage	HIVE-15177 : L'authentification avec Hive échoue lorsque le type d'authentification Kerberos est défini sur et que le principal contient _ fromSubject HOST

Type	Description
Rétroportage	HIVE-14898 : HS2 ne devrait pas enregistrer la pile d'appels pour une erreur d'en-tête d'authentification vide
Rétroportage	HIVE-14493 : Support de partitionnement pour les vues matérialisées
Rétroportage	HIVE-14431 : Reconnaître comme COALESCE CASE
Rétroportage	HIVE-13457 : Création de HS2 REST API points de terminaison pour la surveillance des informations
Rétroportage	HIVE-12342 : Définissez la valeur par défaut de <code>hive.optimize.index.filter</code> sur <code>true</code>
Rétroportage	HIVE-10296 : Exception Cast observée lorsque Hive exécute une requête de jointure multiple sur Metastore
Rétroportage	HIVE-6980 : Supprime une table en utilisant le SQL direct

Amazon EMR 6.6.0 - Modifications de configuration de Hive

- Dans le cadre de la OSS modification [HIVE-20703](#), la propriété permettant de trier les partitions dynamiques `hive.optimize.sort.dynamic.partition` a été remplacée par `hive.optimize.sort.dynamic.partition.threshold`

La configuration `hive.optimize.sort.dynamic.partition.threshold` possède les valeurs potentielles suivantes :

Valeur	Description
0 (default)	Fait de l'optimisation du tri des partitions dynamiques une décision basée sur les coûts lorsque ORC des fichiers sont utilisés. Le nombre maximum d'auteurs autorisés dans les INSERT requêtes est calculé sur la base de (mémoire de l'exécuteur/du conteneur) * (pourcentage de mémoire utilisée par orc) divisé par la mémoire maximale (taille de bande) utilisée par un seul écrivain.
-1	Désactive l'optimisation pour trier complètement les partitions dynamiques.
1	Permet le tri global des partitions dynamiques. Cela permet de maintenir un seul enregistreur ouvert pour chaque valeur de partition dans le réducteur, réduisant ainsi la sollicitation de la mémoire sur les réducteurs.
2 (ou nombre entier supérieur)	Indique à Hive d'utiliser l'entier spécifié comme seuil pour le nombre maximum de rédacteurs.

Amazon EMR 6.6.0 - Problèmes connus liés à Hive

- Les requêtes comportant des fonctions de fenêtrage sur la même colonne que la jointure peuvent entraîner des transformations non valides, comme indiqué dans [HIVE-25278](#), et entraîner des résultats incorrects ou des échecs de requête. Pour contourner le problème, vous pouvez désactiver ces requêtes CBO au niveau de la requête. Contactez AWS le support pour plus d'informations.
- Amazon EMR 6.6.0 inclut la version 3.1.2 du logiciel Hive. [Hive 3.1.2 introduit une fonctionnalité qui divise les fichiers texte s'ils contiennent un en-tête et un pied de page \(-21924\)](#). HIVE L'Apache Tez App Master lit chacun de vos fichiers pour déterminer les points de décalage dans la plage de données. Ces comportements combinés peuvent avoir un impact négatif sur les performances si vos requêtes lisent un grand nombre de petits fichiers texte. Pour contourner le problème, utilisez `CombineHiveInputFormat` et ajustez la taille de division maximale en configurant les propriétés suivantes :

```
SET hive.tez.input.format=org.apache.hadoop.hive.q1.io.CombineHiveInputFormat;
```

```
SET mapreduce.input.fileinputformat.split.maxsize=16777216;
```

- Avec Amazon EMR 6.6.0 à 6.9.x, les INSERT requêtes comportant une partition dynamique et une clause ORDER BY ou SORT BY auront toujours deux réducteurs. Ce problème est dû à la OSS modification [HIVE-20703](#), qui soumet l'optimisation des partitions de tri dynamique à une décision basée sur les coûts. Si votre charge de travail ne nécessite pas le tri des partitions dynamiques, nous vous recommandons de définir la propriété `hive.optimize.sort.dynamic.partition.threshold` sur `-1` pour désactiver la nouvelle fonctionnalité et obtenir le nombre de réducteurs correctement calculé. Ce problème est résolu dans OSS Hive dans le cadre de la version [HIVE-22269](#) et dans Amazon 6.10.0. EMR

Hudi

[Apache Hudi](#) est une infrastructure de gestion des données open source utilisée pour simplifier le traitement des données incrémentielles et le développement de pipelines de données en fournissant des fonctionnalités d'insertion, de mise à jour, de mise à jour/insertion et de suppression au niveau de l'enregistrement. Mise à jour/insertion fait référence à la capacité à insérer des enregistrements dans un ensemble de données existant s'ils n'existent pas déjà ou de les mettre à jour s'ils existent. En gérant efficacement la façon dont les données sont présentées dans Amazon S3, Hudi permet d'ingérer et de mettre à jour des données quasiment en temps réel. Hudi gère soigneusement les métadonnées des actions effectuées sur l'ensemble de données pour faire en sorte que les actions soient atomiques et cohérentes.

Hudi est intégré à [Apache Spark](#), [Apache Hive](#) et [Presto](#). Dans les EMR versions 6.1.0 et ultérieures d'Amazon, Hudi est également intégré à [Trino](#) (Presto). SQL

Avec les EMR versions 5.28.0 et ultérieures d'Amazon, les composants Hudi EMR sont installés par défaut lorsque Spark, Hive, Presto ou Flink sont installés. Vous pouvez utiliser Spark ou l' [DeltaStreamer](#) utilitaire Hudi pour créer ou mettre à jour des ensembles de données Hudi. Vous pouvez utiliser Hive, Spark, Presto ou Flink pour interroger un jeu de données Hudi de manière interactive ou créer des pipelines de traitement de données à l'aide de l'extraction incrémentielle. L'extraction incrémentielle fait référence à la possibilité d'extraire uniquement les données qui ont changé entre deux actions.

Ces fonctionnalités rendent Hudi adapté pour les cas d'utilisation suivants :

- Utilisation des données de streaming provenant de capteurs et d'autres appareils IoT (Internet des objets) nécessitant des événements d'insertion et de mise à jour spécifiques.
- Conformité aux règles de confidentialité des données dans des applications où les utilisateurs peuvent choisir d'être oubliés ou modifier leur consentement quant à la manière dont leurs données peuvent être utilisées.
- Implémentation d'un [système de capture des données de modification \(CDC\)](#) qui vous permet d'appliquer des modifications à un ensemble de données au fil du temps.

Le tableau suivant répertorie la version de Hudi incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Hudi.

Pour la version des composants installés avec Hudi dans cette version, voir la [version 7.2.0 Versions des composants](#).

Informations sur la version Hudi pour emr-7.2.0


Étiquette EMR de sortie Amazon	Version Hudi	Composants installés avec Hudi
emr-7.2.0	Hudi 0.14.1-amzn-1	Not available.

Le tableau suivant répertorie la version de Hudi incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Hudi.

Pour connaître la version des composants installés avec Hudi dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Hudi pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Hudi	Composants installés avec Hudi
emr-6.15.0	Hudi 0.14.0-amzn-0	Not available.

 Note

La EMR version 6.8.0 d'Amazon est fournie avec [Apache Hudi](#) 0.11.1 ; toutefois, les clusters Amazon EMR 6.8.0 sont également compatibles avec l'open source de Hudi 0.12.0. `hudi-spark3.3-bundle_2.12`

Le tableau suivant répertorie la version de Hudi incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Hudi.

Pour la version des composants installés avec Hudi dans cette version, voir la [version 5.36.2 Versions des composants](#).

Informations sur la version Hudi pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Hudi	Composants installés avec Hudi
emr-5.36.2	Hudi 0.10.1-amzn-1	Not available.

Rubriques

- [Comment fonctionne Hudi](#)
- [Considérations et limites relatives à l'utilisation de Hudi sur Amazon EMR](#)
- [Créez un cluster avec Hudi installé](#)
- [Travailler avec un jeu de données Hudi](#)
- [Utilisez le Hudi CLI](#)
- [Historique des sorties de Hudi](#)

Comment fonctionne Hudi

Lorsque vous utilisez Hudi avec AmazonEMR, vous pouvez écrire des données dans l'ensemble de données à l'aide de la source de données Spark API ou de l'utilitaire Hudi DeltaStreamer . Hudi organise un jeu de données dans une structure de répertoire partitionnée sous *basepath* qui est similaire à une table Hive traditionnelle. Les spécificités de la façon dont les données sont présentées en tant que fichiers dans ces répertoires dépendent du type d'ensemble de données que vous choisissez. Vous pouvez choisir CoW (Copy on Write, copie sur écriture) ou MoR (Merge on Read, fusion sur lecture).

Quel que soit le type d'ensemble de données, chaque partition d'un ensemble de données est identifiée de manière unique par son chemin *partitionpath* relatif au chemin *basepath*. Au sein de chaque partition, les enregistrements sont répartis dans plusieurs fichiers de données. Pour de plus amples informations, veuillez consulter [File management](#) dans la documentation Apache Hudi.

Chaque action dans Hudi est associée à un commit correspondant, identifié par un horodatage croissant de façon monotone appelé Instant. Hudi conserve une série de toutes les actions effectuées sur le jeu de données sous forme de chronologie. Hudi s'appuie sur la chronologie pour isoler instantanément les lecteurs et les rédacteurs, et pour permettre de revenir à un point antérieur dans le temps. Pour plus d'informations sur les actions enregistrées par Hudi et l'état des actions, consultez [Timeline](#) dans la documentation Apache Hudi.

Comprendre les types de stockage des jeux de données : copie en écriture ou fusion en lecture

Lorsque vous créez un jeu de données Hudi, vous spécifiez que celui-ci doit être copié en écriture ou fusionné en lecture.

- Copie sur écriture (CoW) – Les données sont stockées dans un format en colonnes (Parquet) et chaque mise à jour crée une nouvelle version des fichiers lors d'une écriture. CoW est le type de stockage par défaut.
- Fusion sur lecture (MoR) – Les données sont stockées en utilisant une combinaison de formats en colonnes (Parquet) et en lignes (Avro). Les mises à jour sont consignées dans des fichiers delta basés sur les lignes et sont compressées si nécessaire pour créer de nouvelles versions des fichiers en colonnes.

Avec les ensembles de données CoW, chaque fois qu'une mise à jour est apportée à un enregistrement, le fichier qui contient l'enregistrement est réécrit avec les valeurs mises à jour. Avec un jeu de données MoR, chaque fois qu'une mise à jour a lieu, Hudi écrit uniquement la ligne du registre modifié. Le type de stockage MoR est mieux adapté aux charges de travail donnant lieu à de nombreuses écritures ou modifications avec moins de lectures. Le type de stockage CoW est mieux adapté aux charges de travail donnant lieu à de nombreuses lectures sur des données qui changent moins souvent.

Hudi fournit trois vues logiques pour l'accès aux données :

- Vue optimisée en lecture – Fournit le dernier ensemble de données validé de tables CoW et le dernier jeu de données compressé de tables MoR.
- Vue incrémentielle : fournit un flux de modifications entre deux actions issues d'un jeu de données CoW pour alimenter les tâches en aval et extraire, transformer, charger (ETL) les flux de travail.
- Vue en temps réel – Fournit les dernières données validées d'une table de MoR en fusionnant en ligne les fichiers de colonnes et les fichiers basés sur des lignes.

Lorsque vous interrogez la vue optimisée en lecture, la requête renvoie toutes les données compressées mais n'inclut pas les dernières validations delta. L'interrogation de ces données fournit de bonnes performances de lecture, mais ignore les données les plus à jour. Lorsque vous interrogez la vue en temps réel, Hudi fusionne les données compressées avec les validations delta sur lecture. Les données les plus à jour sont disponibles pour interrogation, mais la surcharge de calcul de la

fusion rend la requête moins performante. La possibilité d'interroger des données compressées ou des données en temps réel vous permet de choisir entre performances et flexibilité lorsque vous exécutez des requêtes.

Pour plus d'informations sur les avantages et inconvénients des types de stockage, consultez [Storage Types & Views](#) dans la documentation Apache Hudi.

Hudi crée deux tables dans le metastore Hive pour le type de stockage MoR : une table avec le nom que vous avez spécifié, qui est une vue optimisée en lecture, et une table avec le même nom auquel est ajouté `_rt`, qui est une vue en temps réel. Vous pouvez interroger les deux tables.

Enregistrement d'un jeu de données Hudi auprès de votre métastore

Lorsque vous enregistrez une table Hudi dans le métastore Hive, vous pouvez interroger les tables Hudi à l'aide de Hive, Spark SQL ou Presto comme vous le feriez avec n'importe quelle autre table. En outre, vous pouvez intégrer Hudi à AWS Glue en configurant Hive et Spark pour utiliser le catalogue de données AWS Glue comme métastore. Pour les tables MoR, Hudi enregistre le jeu de données sous la forme de deux tables dans le metastore : une table avec le nom que vous avez spécifié, qui est une vue optimisée en lecture, et une table avec le même nom auquel est ajouté `_rt`, qui est une vue en temps réel.

Vous enregistrez une table Hudi auprès du metastore Hive lorsque vous utilisez Spark pour créer un jeu de données Hudi en définissant l'option `HIVE_SYNC_ENABLED_OPT_KEY` sur `"true"` et en fournissant d'autres propriétés requises. Pour plus d'informations, consultez [Travailler avec un jeu de données Hudi](#). En outre, vous pouvez utiliser l'utilitaire de ligne de commande `hive_sync_tool` pour enregistrer un jeu de données Hudi en tant que table dans votre metastore, séparément.

Considérations et limites relatives à l'utilisation de Hudi sur Amazon EMR

- Le champ de clé d'enregistrement ne peut pas être nul ou vide – Le champ que vous spécifiez comme champ de clé d'enregistrement ne peut pas avoir des valeurs `null` ou vides.
- Schéma mis à jour par défaut lors de l'insertion et de l'insertion — Hudi fournit une interface `HoodieRecordPayload` qui détermine comment le jeu de données Hudi en entrée `DataFrame` et le jeu de données Hudi existant sont fusionnés pour produire un nouveau jeu de données mis à jour. Hudi fournit une implémentation par défaut de cette classe `OverwriteWithLatestAvroPayload`, qui remplace les enregistrements existants

et met à jour le schéma tel que spécifié dans l'entrée. DataFrame Pour personnaliser cette logique en vue d'une implémentation de la fusion et des mises à jour partielles, vous pouvez fournir une implémentation de l'interface `HoodieRecordPayload` à l'aide du paramètre `DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY`.

- La suppression exige un schéma – Lors d'une suppression, vous devez spécifier les champs de clé d'enregistrement, de clé de partition et de clé de pré-combinaison. Les autres colonnes peuvent être `null` ou vides, mais le schéma complet est requis.
- Limites des tables MoR – Les tables MoR ne prennent pas en charge le pointage de sauvegarde. Vous pouvez interroger les tables MoR à l'aide de la vue optimisée pour la lecture ou de la vue en temps réel (`tableName_rt`) de SparkSQL, Presto ou Hive. L'utilisation de la vue optimisée pour la lecture expose uniquement les données du fichier de base et n'expose pas une vue fusionnée des données de base et des données de journal.
- Hive
 - Pour l'enregistrement des tables dans le metastore Hive, Hudi s'attend à ce que le serveur Hive Thrift s'exécute sur le port par défaut 10000. Si vous remplacez ce port par un port personnalisé, transmettez l'option `HIVE_URL_OPT_KEY` comme indiqué dans l'exemple suivant.

```
.option(DataSourceWriteOptions.HIVE_URL_OPT_KEY, "jdbc:hive2://localhost:override-port-number
```

- Le type de données `timestamp` dans Spark est enregistré comme type de données `long` dans Hive, et non comme type `timestamp` de Hive.
- Presto
 - Presto ne prend pas en charge la lecture des tables MoR en temps réel dans les versions de Hudi inférieures à 0.6.0.
 - Presto ne prend en charge que les requêtes instantanées.
 - Pour que Presto interprète correctement les colonnes d'un ensemble de données, Hudi définissez la valeur de `hive.parquet_use_column_names` sur `true`.
 - Pour définir la valeur pour une session, dans le shell Presto, exécutez la commande suivante :

```
set session hive.parquet_use_column_names=true
```

- Pour définir la valeur au niveau du cluster, utilisez la classification de configuration `presto-connector-hive` pour définir `hive.parquet.use_column_names` sur `true`, comme illustré dans l'exemple suivant. Pour plus d'informations, consultez [Configuration des applications](#).

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.parquet.use-column-names": "true"
    }
  }
]
```

- HBaseIndice
- La HBase version utilisée pour compiler Hudi peut être différente de celle répertoriée dans le guide de EMR publication. Pour extraire les dépendances correctes pour votre session Spark, exécutez la commande suivante.

```
spark-shell \  
--jars /usr/lib/spark/external/lib/spark-avro.jar,/usr/lib/hudi/cli/lib/*.jar \  
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \  
--conf "spark.sql.hive.convertMetastoreParquet=false"
```

Créez un cluster avec Hudi installé

Avec les EMR versions 5.28.0 et ultérieures d'Amazon, Amazon EMR installe les composants Hudi par défaut lorsque Spark, Hive ou Presto sont installés. Pour utiliser Hudi sur AmazonEMR, créez un cluster avec une ou plusieurs des applications suivantes installées :

- Hadoop
- Hive
- Spark
- Presto
- Flink

Vous pouvez créer un cluster en utilisant le AWS Management Console AWS CLI, le ou Amazon EMRAPI.

Pour créer un cluster avec Hudi à l'aide du AWS Management Console

1. Accédez à la nouvelle EMR console Amazon et sélectionnez Basculer vers l'ancienne console dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).
2. Choisissez Créer un cluster et Go to advanced options (Aller aux options avancées).
3. Sous Software Configuration (Configuration logicielle), choisissez emr-5.28.0 ou version ultérieure pour Release (Version) et sélectionnez Hadoop, Hive, Spark, Presto et Tez avec les autres applications dont votre cluster a besoin.
4. Configurez d'autres options si nécessaire pour votre application, puis choisissez Next (Suivant).
5. Configurez les options pour Hardware (Matériel) and General cluster settings (Paramètres de cluster généraux) à votre guise.
6. Pour les options de sécurité, nous vous recommandons de sélectionner une paire de EC2 clés que vous pouvez utiliser pour vous connecter à la ligne de commande du nœud principal à l'aide deSSH. Cela vous permet d'exécuter les commandes Spark shell, les commandes Hive et CLI les CLI commandes Hudi décrites dans ce guide.
7. Configurez les autres options de sécurité comme vous le souhaitez et choisissez Create cluster (Créer un cluster).

Travailler avec un jeu de données Hudi

Hudi prend en charge l'insertion, la mise à jour et la suppression de données dans des jeux de données Hudi via Spark. Pour plus d'informations, consultez [Writing Hudi tables](#) dans la documentation Apache Hudi.

Les exemples suivants montrent comment lancer le shell interactif Spark, utiliser Spark submit ou utiliser Amazon EMR Notebooks pour travailler avec Hudi sur Amazon. EMR Vous pouvez également utiliser l' DeltaStreamer utilitaire Hudi ou d'autres outils pour écrire dans un ensemble de données. Tout au long de cette section, les exemples montrent comment utiliser des ensembles de données à l'aide du shell Spark lorsque vous êtes connecté au nœud principal en SSH tant qu'hadooputilisateur par défaut.

Lancez le shell Spark à l'aide d'Amazon EMR 6.7 et versions ultérieures

Lors de l'exécution `spark-shell` ou de `spark-sql` l'utilisation d'Amazon EMR 6.7.0 ou version ultérieure, passez les commandes suivantes. `spark-submit`

Note

Amazon EMR 6.7.0 utilise [Apache Hudi 0.11.0-amzn-0](#), qui contient des améliorations significatives par rapport aux versions précédentes de Hudi. Pour plus d'informations, consultez le [Guide de migration Apache Hudi 0.11.0](#). Les exemples de cet onglet reflètent ces changements.

Pour ouvrir le shell Spark sur le nœud primaire

1. Connectez-vous au nœud principal à l'aide de SSH. Pour plus d'informations, consultez [Se connecter au nœud principal SSH à l'aide](#) de l'Amazon EMR Management Guide.
2. Entrez la commande suivante pour lancer le shell Spark. Pour utiliser la PySpark coque, remplacez *spark-shell* avec *pyspark*.

```
spark-shell --jars /usr/lib/hudi/hudi-spark-bundle.jar \  
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \  
--conf  
"spark.sql.catalog.spark_catalog=org.apache.spark.sql.hudi.catalog.HoodieCatalog"  
\  
--conf "spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
```

Lancez le shell Spark à l'aide d'Amazon EMR 6.6 et versions antérieures

Lorsque vous exécutez `spark-shell` ou `spark-sql` utilisez Amazon EMR 6.6.x ou une version antérieure, passez les commandes suivantes. `spark-submit`

Note

- Amazon EMR 6.2, 5.31 et versions ultérieures (Hudi 0.6.x et versions ultérieures) peuvent les omettre de la configuration. `spark-avro.jar`
- Amazon EMR 6.5, 5.35 et versions ultérieures (Hudi 0.9.x et versions ultérieures) peuvent ne pas figurer dans la configuration. `spark.sql.hive.convertMetastoreParquet=false`
- Amazon EMR 6.6, 5.36 et versions ultérieures (Hudi 0.10.x et versions ultérieures) doivent inclure la `HoodieSparkSessionExtension` configuration telle que décrite dans le guide Spark de la [version : 0.10.0](#) :

```
--conf
"spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
\  

```

Pour ouvrir le shell Spark sur le nœud primaire

1. Connectez-vous au nœud principal à l'aide de SSH. Pour plus d'informations, consultez [Se connecter au nœud principal SSH à l'aide](#) de l'Amazon EMR Management Guide.
2. Entrez la commande suivante pour lancer le shell Spark. Pour utiliser la PySpark coque, remplacez *spark-shell* avec *pyspark*.

```
spark-shell \  
--conf "spark.serializer=org.apache.spark.serializer.KryoSerializer" \  
--conf "spark.sql.hive.convertMetastoreParquet=false" \  
--jars /usr/lib/hudi/hudi-spark-bundle.jar,/usr/lib/spark/external/lib/spark-  
avro.jar
```

Utiliser Hudi avec Amazon EMR Notebooks utilisant Amazon EMR 6.7 et versions ultérieures

Pour utiliser Hudi avec Amazon EMR Notebooks, vous devez d'abord copier les fichiers JAR Hudi depuis le système de fichiers local vers HDFS le nœud principal du cluster de blocs-notes. Vous utilisez ensuite l'éditeur de bloc-notes pour configurer votre EMR bloc-notes afin qu'il utilise Hudi.

Pour utiliser Hudi avec Amazon Notebooks EMR

1. Créez et lancez un cluster pour Amazon EMR Notebooks. Pour plus d'informations, consultez la section [Création de EMR clusters Amazon pour ordinateurs portables](#) dans le guide EMR de gestion Amazon.
2. Connectez-vous au nœud principal du cluster à l'aide SSH des fichiers jar du système de fichiers local, puis copiez-les HDFS comme indiqué dans les exemples suivants. Dans l'exemple, nous créons un répertoire dans HDFS pour faciliter la gestion des fichiers. Vous pouvez choisir votre propre destination HDFS, si vous le souhaitez.

```
hdfs dfs -mkdir -p /apps/hudi/lib
```

```
hdfs dfs -copyFromLocal /usr/lib/hudi/hudi-spark-bundle.jar /apps/hudi/lib/hudi-spark-bundle.jar
```

3. Ouvrez l'éditeur de bloc-notes, entrez le code de l'exemple suivant et exécutez-le.

```
%%configure
{ "conf": {
    "spark.jars":"hdfs:///apps/hudi/lib/hudi-spark-bundle.jar",
    "spark.serializer":"org.apache.spark.serializer.KryoSerializer",
    "spark.sql.catalog.spark_catalog":
    "org.apache.spark.sql.hudi.catalog.HoodieCatalog",

    "spark.sql.extensions":"org.apache.spark.sql.hudi.HoodieSparkSessionExtension"
}}
```

Utiliser Hudi avec Amazon EMR Notebooks utilisant Amazon EMR 6.6 et versions antérieures

Pour utiliser Hudi avec Amazon EMR Notebooks, vous devez d'abord copier les fichiers JAR Hudi depuis le système de fichiers local vers HDFS le nœud principal du cluster de blocs-notes. Vous utilisez ensuite l'éditeur de bloc-notes pour configurer votre EMR bloc-notes afin qu'il utilise Hudi.

Pour utiliser Hudi avec Amazon Notebooks EMR

1. Créez et lancez un cluster pour Amazon EMR Notebooks. Pour plus d'informations, consultez la section [Création de EMR clusters Amazon pour ordinateurs portables](#) dans le guide EMR de gestion Amazon.
2. Connectez-vous au nœud principal du cluster à l'aide SSH des fichiers jar du système de fichiers local, puis copiez-les HDFS comme indiqué dans les exemples suivants. Dans l'exemple, nous créons un répertoire dans HDFS pour faciliter la gestion des fichiers. Vous pouvez choisir votre propre destination HDFS, si vous le souhaitez.

```
hdfs dfs -mkdir -p /apps/hudi/lib
```

```
hdfs dfs -copyFromLocal /usr/lib/hudi/hudi-spark-bundle.jar /apps/hudi/lib/hudi-spark-bundle.jar
```

```
hdfs dfs -copyFromLocal /usr/lib/spark/external/lib/spark-avro.jar /apps/hudi/lib/spark-avro.jar
```

3. Ouvrez l'éditeur de bloc-notes, entrez le code de l'exemple suivant et exécutez-le.

```
{ "conf": {  
    "spark.jars": "hdfs:///apps/hudi/lib/hudi-spark-bundle.jar,hdfs:///apps/hudi/lib/spark-avro.jar",  
    "spark.serializer": "org.apache.spark.serializer.KryoSerializer",  
    "spark.sql.hive.convertMetastoreParquet": "false"  
}}
```

Initialisation d'une session Spark pour Hudi

Lorsque vous utilisez Scala, vous devez importer les classes suivantes dans votre session Spark. Vous devez le faire une fois par session Spark.

```
import org.apache.spark.sql.SaveMode  
import org.apache.spark.sql.functions._  
import org.apache.hudi.DataSourceWriteOptions  
import org.apache.hudi.DataSourceReadOptions  
import org.apache.hudi.config.HoodieWriteConfig  
import org.apache.hudi.hive.MultiPartKeyValueExtractor  
import org.apache.hudi.hive.HiveSyncConfig  
import org.apache.hudi.sync.common.HoodieSyncConfig
```

Écrire dans un jeu de données Hudi

Les exemples suivants montrent comment créer un jeu de données Hudi DataFrame et l'écrire sous forme de jeu de données Hudi.

Note

Pour coller des exemples de code dans le shell Spark, tapez **:paste** à l'invite, collez l'exemple, puis appuyez sur **CTRL + D**.

Chaque fois que vous écrivez un dans DataFrame un jeu de données Hudi, vous devez spécifier `DataSourceWriteOptions`. Beaucoup de ces options sont susceptibles d'être identiques

dans les opérations d'écriture. L'exemple suivant spécifie les options communes à l'aide de la variable `hoodieOptions`, que les exemples suivants utilisent.

Écrire à l'aide de Scala avec Amazon EMR 6.7 et versions ultérieures

Note

Amazon EMR 6.7.0 utilise [Apache Hudi 0.11.0-amzn-0](#), qui contient des améliorations significatives par rapport aux versions précédentes de Hudi. Pour plus d'informations, consultez le [Guide de migration Apache Hudi 0.11.0](#). Les exemples de cet onglet reflètent ces changements.

```
// Create a DataFrame
val inputDF = Seq(
  ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
  ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
  ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
  ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
  ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
  ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z")
).toDF("id", "creation_date", "last_update_time")

//Specify common DataSourceWriteOptions in the single hoodieOptions variable
val hoodieOptions = Map[String,String](
  HoodieWriteConfig.TBL_NAME.key -> "tableName",
  DataSourceWriteOptions.TABLE_TYPE.key -> "COPY_ON_WRITE",
  DataSourceWriteOptions.RECORDKEY_FIELD_OPT_KEY -> "id",
  DataSourceWriteOptions.PARTITIONPATH_FIELD_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.PRECOMBINE_FIELD_OPT_KEY -> "last_update_time",
  DataSourceWriteOptions.HIVE_SYNC_ENABLED_OPT_KEY -> "true",
  DataSourceWriteOptions.HIVE_TABLE_OPT_KEY -> "tableName",
  DataSourceWriteOptions.HIVE_PARTITION_FIELDS_OPT_KEY -> "creation_date",
  HoodieSyncConfig.META_SYNC_PARTITION_EXTRACTOR_CLASS.key ->
  "org.apache.hudi.hive.MultiPartKeyValueExtractor",
  HoodieSyncConfig.META_SYNC_ENABLED.key -> "true",
  HiveSyncConfig.HIVE_SYNC_MODE.key -> "hms",
  HoodieSyncConfig.META_SYNC_TABLE_NAME.key -> "tableName",
  HoodieSyncConfig.META_SYNC_PARTITION_FIELDS.key -> "creation_date"
)

// Write the DataFrame as a Hudi dataset
```

```
(inputDF.write
  .format("hudi")
  .options(hudiOptions)
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,"insert")
  .mode(SaveMode.Overwrite)
  .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

Écrire à l'aide de Scala avec Amazon EMR 6.6 et versions antérieures

```
// Create a DataFrame
val inputDF = Seq(
  ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
  ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
  ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
  ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
  ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
  ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z")
).toDF("id", "creation_date", "last_update_time")

//Specify common DataSourceWriteOptions in the single hudiOptions variable
val hudiOptions = Map[String,String](
  HoodieWriteConfig.TABLE_NAME -> "tableName",
  DataSourceWriteOptions.TABLE_TYPE_OPT_KEY -> "COPY_ON_WRITE",
  DataSourceWriteOptions.RECORDKEY_FIELD_OPT_KEY -> "id",
  DataSourceWriteOptions.PARTITIONPATH_FIELD_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.PRECOMBINE_FIELD_OPT_KEY -> "last_update_time",
  DataSourceWriteOptions.HIVE_SYNC_ENABLED_OPT_KEY -> "true",
  DataSourceWriteOptions.HIVE_TABLE_OPT_KEY -> "tableName",
  DataSourceWriteOptions.HIVE_PARTITION_FIELDS_OPT_KEY -> "creation_date",
  DataSourceWriteOptions.HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY ->
  classOf[MultiPartKeysValueExtractor].getName
)

// Write the DataFrame as a Hudi dataset
(inputDF.write
  .format("org.apache.hudi")
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
DataSourceWriteOptions.INSERT_OPERATION_OPT_VAL)
  .options(hudiOptions)
  .mode(SaveMode.Overwrite)
  .save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

Écrivez en utilisant PySpark

```
# Create a DataFrame
inputDF = spark.createDataFrame(
    [
        ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
        ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
        ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
        ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z"),
        ("104", "2015-01-02", "2015-01-01T12:15:00.512679Z"),
        ("105", "2015-01-02", "2015-01-01T13:51:42.248818Z"),
    ],
    ["id", "creation_date", "last_update_time"]
)

# Specify common DataSourceWriteOptions in the single hudiOptions variable
hudiOptions = {
    'hoodie.table.name': 'tableName',
    'hoodie.datasource.write.recordkey.field': 'id',
    'hoodie.datasource.write.partitionpath.field': 'creation_date',
    'hoodie.datasource.write.precombine.field': 'last_update_time',
    'hoodie.datasource.hive_sync.enable': 'true',
    'hoodie.datasource.hive_sync.table': 'tableName',
    'hoodie.datasource.hive_sync.partition_fields': 'creation_date',
    'hoodie.datasource.hive_sync.partition_extractor_class':
    'org.apache.hudi.hive.MultiPartKeysValueExtractor'
}

# Write a DataFrame as a Hudi dataset
inputDF.write \
    .format('org.apache.hudi') \
    .option('hoodie.datasource.write.operation', 'insert') \
    .options(**hudiOptions) \
    .mode('overwrite') \
    .save('s3://DOC-EXAMPLE-BUCKET/myhudidataset/')
```

Note

Vous pouvez voir « hoodie » au lieu de Hudi dans les exemples de code et les notifications. La base de code Hudi utilise largement l'ancienne orthographe « hoodie ».

DataSourceWriteOptions référence pour Hudi

Option	Description
TABLE_NAME	Nom de la table sous laquelle enregistrer l'ensemble de données.
TABLE_TYPE_OPT_KEY	Facultatif. Spécifie si l'ensemble de données est créé en tant que "COPY_ON_WRITE" ou "MERGE_ON_READ". L'argument par défaut est "COPY_ON_WRITE".
RECORDKEY_FIELD_OPT_KEY	Champ de clé d'enregistrement dont la valeur sera utilisée en tant que composant <code>recordKey</code> de <code>HoodieKey</code> . La valeur réelle sera obtenue en appelant <code>.toString()</code> sur la valeur de champ. Des champs imbriqués peuvent être spécifiés à l'aide de la notation par points, par exemple, <code>a.b.c</code> .
PARTITIONPATH_FIELD_OPT_KEY	Champ de chemin de partition dont la valeur sera utilisée en tant que composant <code>partitionPath</code> de <code>HoodieKey</code> . La valeur réelle sera obtenue en appelant <code>.toString()</code> sur la valeur de champ.
PRECOMBINE_FIELD_OPT_KEY	Champ utilisé dans la pré-combinaison avant l'écriture réelle. Lorsque deux enregistrements ont la même valeur de clé, Hudi sélectionne celui avec la plus grande valeur pour le champ de pré-combinaison, comme déterminé par <code>Object.compareTo(...)</code> .

Les options suivantes sont nécessaires uniquement pour enregistrer la table du jeu de données Hudi dans votre metastore. Si vous n'enregistrez pas votre jeu de données Hudi en tant que table dans le metastore Hive, ces options ne sont pas requises.

DataSourceWriteOptions référence pour Hive

Option	Description
HIVE_DATABASE_OPT_KEY	Base de données Hive avec laquelle synchroniser. L'argument par défaut est "default" .
HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY	Classe utilisée pour extraire les valeurs de champ de partition dans les colonnes de partition Hive.
HIVE_PARTITION_FIELDS_OPT_KEY	Champ dans l'ensemble de données à utiliser pour déterminer les colonnes de partition Hive.
HIVE_SYNC_ENABLED_OPT_KEY	Lorsqu'elle est définie sur "true", enregistre l'ensemble de données auprès du metastore Apache Hive. L'argument par défaut est "false".
HIVE_TABLE_OPT_KEY	Obligatoire. Nom de la table dans Hive avec laquelle synchroniser. Par exemple, "my_hudi_table_cow" .
HIVE_USER_OPT_KEY	Facultatif. Nom d'utilisateur Hive à utiliser lors de la synchronisation. Par exemple, "hadoop".
HIVE_PASS_OPT_KEY	Facultatif. Mot de passe Hive pour l'utilisateur spécifié par HIVE_USER_OPT_KEY .
HIVE_URL_OPT_KEY	Le URL métastore Hive.

Données d'Upsert

L'exemple suivant montre comment insérer des données en écrivant un DataFrame.

Contrairement à l'exemple d'insertion précédent, la valeur OPERATION_OPT_KEY est définie sur UPSERT_OPERATION_OPT_VAL. En outre, `.mode(SaveMode.Append)` est spécifié pour indiquer que l'enregistrement doit être ajouté.

Upsert à l'aide de Scala avec Amazon EMR 6.7 et versions ultérieures

Note

Amazon EMR 6.7.0 utilise [Apache Hudi 0.11.0-amzn-0](#), qui contient des améliorations significatives par rapport aux versions précédentes de Hudi. Pour plus d'informations, consultez le [Guide de migration Apache Hudi 0.11.0](#). Les exemples de cet onglet reflètent ces changements.

```
// Create a new DataFrame from the first row of inputDF with a different creation_date
value
val updateDF = inputDF.limit(1).withColumn("creation_date", lit("new_value"))

(updateDF.write
  .format("hudi")
  .options(hudiOptions)
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "upsert")
  .mode(SaveMode.Append)
  .save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

Upsert en utilisant Scala avec Amazon EMR 6.6 et versions antérieures

```
// Create a new DataFrame from the first row of inputDF with a different creation_date
value
val updateDF = inputDF.limit(1).withColumn("creation_date", lit("new_value"))

(updateDF.write
  .format("org.apache.hudi")
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
DataSourceWriteOptions.UPSERT_OPERATION_OPT_VAL)
  .options(hudiOptions)
  .mode(SaveMode.Append)
  .save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

Upsert en utilisant PySpark

```
from pyspark.sql.functions import lit
```

```
# Create a new DataFrame from the first row of inputDF with a different creation_date
value
updateDF = inputDF.limit(1).withColumn('creation_date', lit('new_value'))

updateDF.write \
    .format('org.apache.hudi') \
    .option('hoodie.datasource.write.operation', 'upsert') \
    .options(**hoodieOptions) \
    .mode('append') \
    .save('s3://DOC-EXAMPLE-BUCKET/myhuidataset/')
```

Supprimer un enregistrement

Pour supprimer définitivement un enregistrement, vous pouvez insérer une charge utile vide. Dans ce cas, l'option `PAYLOAD_CLASS_OPT_KEY` spécifie la classe `EmptyHoodieRecordPayload`. L'exemple utilise le même `DataFrame`, `updateDF`, utilisé dans l'exemple `upsert` pour spécifier le même enregistrement.

Supprimer à l'aide de Scala avec Amazon EMR 6.7 et versions ultérieures

Note

Amazon EMR 6.7.0 utilise [Apache Hudi 0.11.0-amzn-0](#), qui contient des améliorations significatives par rapport aux versions précédentes de Hudi. Pour plus d'informations, consultez le [Guide de migration Apache Hudi 0.11.0](#). Les exemples de cet onglet reflètent ces changements.

```
(updateDF.write
    .format("hudi")
    .options(hudiOptions)
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "delete")
    .mode(SaveMode.Append)
    .save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

Supprimer à l'aide de Scala avec Amazon EMR 6.6 et versions antérieures

```
(updateDF.write
    .format("org.apache.hudi")
    .option(DataSourceWriteOptions.OPERATION_OPT_KEY,
DataSourceWriteOptions.UPSERT_OPERATION_OPT_VAL)
```

```
.option(DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY,
"org.apache.hudi.common.model.EmptyHoodieRecordPayload")
.mode(SaveMode.Append)
.save("s3://DOC-EXAMPLE-BUCKET/myhudidataset/"))
```

Supprimer en utilisant PySpark

```
updateDF.write \
    .format('org.apache.hudi') \
    .option('hoodie.datasource.write.operation', 'upsert') \
    .option('hoodie.datasource.write.payload.class',
'org.apache.hudi.common.model.EmptyHoodieRecordPayload') \
    .options(**hudiOptions) \
    .mode('append') \
    .save('s3://DOC-EXAMPLE-BUCKET/myhudidataset/')
```

Vous pouvez également supprimer définitivement des données en définissant `OPERATION_OPT_KEY` sur `DELETE_OPERATION_OPT_VAL` pour supprimer tous les enregistrements de le jeu de données que vous soumettez. Pour obtenir des instructions sur les suppressions logicielles et pour plus d'informations sur la suppression de données stockées dans des tables Hudi, consultez [Suppressions](#) dans la documentation d'Apache Hudi.

Lire à partir d'un jeu de données Hudi

Pour récupérer les données à l'heure actuelle, Hudi effectue des requêtes instantanées par défaut. Voici un exemple d'interrogation de le jeu de données écrit dans S3 dans [Écrire dans un jeu de données Hudi](#). Remplacez `s3://DOC-EXAMPLE-BUCKET/myhudidataset` avec le chemin de votre table, et ajoutez des astérisques génériques pour chaque niveau de partition, plus un astérisque supplémentaire. Dans cet exemple, il existe un niveau de partition. Nous avons donc ajouté deux symboles génériques.

Lisez [Utiliser Scala avec Amazon EMR 6.7 et versions ultérieures](#)

Note

Amazon EMR 6.7.0 utilise [Apache Hudi 0.11.0-amzn-0](#), qui contient des améliorations significatives par rapport aux versions précédentes de Hudi. Pour plus d'informations, consultez le [Guide de migration Apache Hudi 0.11.0](#). Les exemples de cet onglet reflètent ces changements.

```
val snapshotQueryDF = spark.read
  .format("hudi")
  .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset")
  .show()
```

Lisez Utiliser Scala avec Amazon EMR 6.6 et versions antérieures

```
(val snapshotQueryDF = spark.read
  .format("org.apache.hudi")
  .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset" + "/*/*"))

snapshotQueryDF.show()
```

Lisez en utilisant PySpark

```
snapshotQueryDF = spark.read \
  .format('org.apache.hudi') \
  .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset' + '/*/*')

snapshotQueryDF.show()
```

Requêtes incrémentielles

Vous pouvez également effectuer des requêtes incrémentielles avec Hudi pour obtenir un flux d'enregistrements modifiés depuis un horodatage de validation donné. Pour ce faire, définissez le champ `QUERY_TYPE_OPT_KEY` sur `QUERY_TYPE_INCREMENTAL_OPT_VAL`. Ajoutez ensuite une valeur pour `BEGIN_INSTANTTIME_OPT_KEY` pour obtenir tous les enregistrements écrits depuis l'heure spécifiée. Les requêtes incrémentielles sont généralement dix fois plus efficaces que leurs homologues par lots, car elles ne traitent que les enregistrements modifiés.

Lorsque vous effectuez des requêtes incrémentielles, utilisez le chemin de la table racine (de base) sans les astérisques génériques utilisés pour les requêtes Snapshot.

Note

Presto ne prend pas en charge les requêtes incrémentielles.

Requêtes incrémentielles à l'aide de Scala

```
(val incQueryDF = spark.read
  .format("org.apache.hudi")
  .option(DataSourceReadOptions.QUERY_TYPE_OPT_KEY,
    DataSourceReadOptions.QUERY_TYPE_INCREMENTAL_OPT_VAL)
  .option(DataSourceReadOptions.BEGIN_INSTANTTIME_OPT_KEY, <beginInstantTime>)
  .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset" ))

incQueryDF.show()
```

Requêtes incrémentielles utilisant PySpark

```
readOptions = {
  'hoodie.datasource.query.type': 'incremental',
  'hoodie.datasource.read.begin.instanttime': <beginInstantTime>,
}

incQueryDF = spark.read \
  .format('org.apache.hudi') \
  .options(**readOptions) \
  .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset')

incQueryDF.show()
```

Pour plus d'informations sur la lecture d'jeux de données Hudi, consultez la rubrique [Interrogation de tables Hudi](#) dans la documentation d'Apache Hudi.

Utilisez le Hudi CLI

Vous pouvez utiliser le Hudi CLI pour administrer des ensembles de données Hudi afin d'afficher des informations sur les validations, le système de fichiers, les statistiques, etc. Vous pouvez également utiliser le CLI pour effectuer manuellement des compactages, planifier des compactages ou annuler des compactages planifiés. Pour plus d'informations, consultez [Interaction via CLI](#) dans la documentation d'Apache Hudi.

Pour démarrer le Hudi CLI et se connecter à un jeu de données

1. Connectez-vous au nœud principal à l'aide de SSH. Pour plus d'informations, consultez [Connect to the master node SSH à l'aide](#) de l'Amazon EMR Management Guide.

- Sur la ligne de commande, tapez `/usr/lib/hudi/cli/bin/hudi-cli.sh`.

L'invite de commande devient `hudi->`.

- Tapez le code suivant pour vous connecter à un ensemble de données. Remplacez `s3://DOC-EXAMPLE-BUCKET/myhudidataset` avec le chemin d'accès au jeu de données avec lequel vous souhaitez travailler. La valeur que nous utilisons est la même que la valeur établie dans les exemples précédents.

```
connect --path s3://DOC-EXAMPLE-BUCKET/myhudidataset
```

L'invite de commandes change pour inclure l'ensemble de données auquel vous êtes connecté, comme illustré dans l'exemple suivant.

```
hudi:myhudidataset->
```

Historique des sorties de Hudi

Le tableau suivant répertorie la version de Hudi incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section [Version des composants de votre version](#) dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Hudi

Étiquette Amazon EMR Release	Version Hudi	Composants installés avec Hudi
emr-7.2.0	0,14.1-amzn-1	Not available.
emr-5.36.2	0.10.1-amzn-1	Not available.
emr-7.1.0	0.14.1-amzn-0	Not available.
emr-7.0.0	0.14.0-amzn-1	Not available.
emr-6.15.0	0.14.0-amzn-0	Not available.

Étiquette Amazon EMR Release	Version Hudi	Composants installés avec Hudi
emr-6.14.0	0.13.1-amzn-2	Not available.
emr-6.13.0	0.13.1-amzn-1	Not available.
emr-6.12.0	0.13.1-amzn-0	Not available.
emr-6.11.1	0.13.0-amzn-0	Not available.
emr-6.11.0	0.13.0-amzn-0	Not available.
emr-6.10.1	0.12.2-amzn-0	Not available.
emr-6.10.0	0.12.2-amzn-0	Not available.
emr-6.9.1	0.12.1-amzn-0	Not available.
emr-6.9.0	0.12.1-amzn-0	Not available.
emr-6.8.1	0.11.1-amzn-0	Not available.
emr-6.8.0	0.11.1-amzn-0	Not available.
emr-6.7.0	0.11.0-amzn-0	Not available.
emr-5.36.1	0.10.1-amzn-1	Not available.
emr-5.36.0	0.10.1-amzn-1	Not available.
emr-6.6.0	0.10.1-amzn-0	Not available.
emr-5.35.0	0.9.0-amzn-2	Not available.
emr-6.5.0	0.9.0-amzn-1	Not available.
emr-6.4.0	0.8.0-amzn-0	Not available.
emr-6.3.1	0.7.0-amzn-0	Not available.
emr-6.3.0	0.7.0-amzn-0	Not available.

Étiquette Amazon EMR Release	Version Hudi	Composants installés avec Hudi
emr-6.2.1	0.6.0-amzn-1	Not available.
emr-6.2.0	0.6.0-amzn-1	Not available.
emr-6.1.1	0.5.2-incubating-amzn-2	Not available.
emr-6.1.0	0.5.2-incubating-amzn-2	Not available.
emr-6.0.1	0.5.0-incubating-amzn-1	Not available.
emr-6.0.0	0.5.0-incubating-amzn-1	Not available.
emr-5.34.0	0.9.0-amzn-0	Not available.
emr-5.33.1	0.7.0-amzn-1	Not available.
emr-5.33.0	0.7.0-amzn-1	Not available.
emr-5.32.1	0.6.0-amzn-0	Not available.
emr-5.32.0	0.6.0-amzn-0	Not available.
emr-5.31.1	0.6.0-amzn-0	Not available.
emr-5.31.0	0.6.0-amzn-0	Not available.
emr-5.30.2	0.5.2-incubating	Not available.
emr-5.30.1	0.5.2-incubating	Not available.
emr-5.30.0	0.5.2-incubating	Not available.
mr-5.29.0	0.5.0-incubating	Not available.
emr-5.28.1	0.5.0-incubating	Not available.
emr-5.28.0	0.5.0-incubating	Not available.

Hue

Hue (Hadoop User Experience) est une interface utilisateur graphique open source, basée sur le Web, à utiliser avec Amazon EMR et Apache Hadoop. Hue regroupe plusieurs projets d'écosystème Hadoop différents au sein d'une interface configurable. Amazon EMR a également ajouté des personnalisations spécifiques à Hue sur AmazonEMR. Hue joue le rôle de serveur frontal pour les applications exécutées sur votre cluster, ce qui vous permet d'interagir avec les applications via une interface plus conviviale ou qui vous est plus familière. Grâce aux applications figurant dans Hue, telles que les éditeurs Hive et Pig, il n'est plus nécessaire de se connecter au cluster pour exécuter des scripts de façon interactive avec le shell de chaque application. Une fois que le cluster est lancé, vous pouvez interagir entièrement avec vos applications en utilisant Hue ou une interface similaire. Pour plus d'informations sur Hue, consultez <http://gethue.com>.

Hue est installé par défaut lorsque vous lancez votre cluster à l'aide de la EMR console Amazon. Vous pouvez choisir de ne pas installer Hue en utilisant les options avancées de la EMR console Amazon lorsque vous lancez un cluster, ou en spécifiant explicitement `--applicationsoption`, et en omettant Hue, lorsque vous l'utilisez `create-cluster` depuis le AWS CLI.

Rubriques

- [Informations sur la version de Hue](#)
- [Fonctionnalités prises en charge et non prises en charge de Hue sur Amazon EMR](#)
- [Considérations](#)
- [Connexion à l'interface utilisateur Web de Hue](#)
- [Utilisation de Hue avec une base de données distante sur Amazon RDS](#)
- [Configurations avancées pour Hue](#)
- [Historique des versions de Hue](#)

Informations sur la version de Hue

Version Hue pour 7.2.0

Le tableau suivant répertorie la version de Hue incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Hue.

Pour la version des composants installés avec Hue dans cette version, voir [Versions des composants de la version 7.2.0](#).

Informations de version Hue pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Hue	Composants installés avec Hue
emr-7.2.0	Hue 4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Version de Hue pour 6.15.0

Le tableau suivant répertorie la version de Hue incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Hue.

Pour connaître la version des composants installés avec Hue dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Hue pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Hue	Composants installés avec Hue
emr-6.15.0	Hue 4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Étiquette EMR de sortie Amazon	Version Hue	Composants installés avec Hue
		timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Version Hue pour 5.36.2

Le tableau suivant répertorie la version de Hue incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Hue.

Pour la version des composants installés avec Hue dans cette version, voir la [version 5.36.2 Versions des composants](#).

Informations de version Hue pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Hue	Composants installés avec Hue
emr-5.36.2	Hue 4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Fonctionnalités prises en charge et non prises en charge de Hue sur Amazon EMR

Avec Amazon EMR 7.0.0 et versions ultérieures, Hue nécessite Python 3.9 ou supérieur. Lorsque vous utilisez le chiffrement en transit, vous devez utiliser une clé de chiffrement de 2 048 bits.

- Amazon S3 et Hadoop File System (HDFS) Browser
 - Avec les autorisations appropriées, vous pouvez parcourir et déplacer des données entre le HDFS stockage éphémère et les compartiments S3 appartenant à votre compte.
 - Par défaut, les superutilisateurs de Hue peuvent accéder à tous les fichiers auxquels les EMR IAM rôles Amazon sont autorisés à accéder. Les utilisateurs nouvellement créés n'ont pas automatiquement les autorisations d'accéder à l'explorateur de fichiers Amazon S3 et doivent avoir les autorisations `filebrowser.s3_access` activées pour leur groupe.
- Hive – Exécutez des requêtes interactives sur vos données. Vous pouvez également créer un prototype d'interrogation par programmation ou par lot.
- Pig – Exécutez des scripts sur vos données ou exécutez des commandes interactives.
- Oozie – Créer et surveiller des flux de travail Oozie.
- Gestionnaire Metastore – Affichez et manipulez le contenu de la metastore Hive (importation/ création, suppression, etc.).
- Explorateur de tâches – Affichez l'état des tâches Hadoop que vous avez soumises.
- Gestion des utilisateurs — Gérez les comptes Hue et intégrez LDAP les utilisateurs à Hue.
- AWS Exemples — Il existe plusieurs « ready-to-run » exemples de traitement d'échantillons de données provenant de divers AWS services à l'aide d'applications dans Hue. Lorsque vous vous connectez à Hue, vous êtes dirigé vers l'application Hue Home où les exemples de données sont préinstallés.
- Livy Server n'est pris en charge que dans les EMR versions 5.9.0 et ultérieures d'Amazon.
- Pour utiliser Hue Notebook pour Spark, vous devez installer Hue avec Livy et Spark.
- Trino/Presto – Hue prend en charge la connexion aux serveurs de données Trino et Presto. Avec Amazon EMR 7.0.0 et versions ultérieures, cette connectivité est facilitée par `trino-python-client` des `presto-python-client` connecteurs. Les versions inférieures à 7.0.0 utilisent un JDBC connecteur via une JDBC passerelle.
- Le tableau de bord Hue n'est pas pris en charge.
- Postgre n'SQLest pas pris en charge.

Considérations

Tenez compte des limites suivantes lorsque vous utilisez [Hue](#) sur AmazonEMR.

Performances avec de grandes tables de métadonnées Hue

Si la base de données de métadonnées Hue devient trop volumineuse, les performances risquent de se dégrader. Pour vérifier la taille des tables, connectez-vous d'abord au nœud principal du EC2 cluster Amazon EMR sur Amazon SSH et exécutez la commande `sudo mysql -u root` pour démarrer le My SQLCLI. Pour obtenir la taille de votre table, exécutez la requête `SELECT COUNT(*) FROM hue.<table_name>`. Consultez ce qui suit pour quoi `<table_name>` peut être :

- document_bureau
- document_bureau 2
- oozie_job
- requête beeswax_saved
- séance de cire d'abeille
- histoire de beeswax_query

Si l'exécution de cette requête renvoie un nombre supérieur à 100 000, vous devez exécuter la commande de nettoyage suivante pour supprimer les anciens enregistrements.

```
cd /opt/cloudera/parcels/CDH/lib/hue # Hue home directory ./build/env/bin/hue
desktop_document_cleanup
```

Pour plus d'informations sur le nettoyage de votre base de données, consultez la [page de référence de la documentation Hue](#).

Hue ne nettoie pas automatiquement les tables, mais les EMR versions 5.12.0 et ultérieures d'Amazon proposent une méthode pour supprimer les anciens documents des tables. Créez le script shell suivant et exécutez-le en tant qu'étape dans un EMR cluster Amazon avec un paramètre entier qui représente le nombre maximal de jours pendant lesquels les documents doivent être conservés dans la base de données de métadonnées.

```
#!/bin/bash
if grep isMaster /mnt/var/lib/info/instance.json | grep false;
then
    echo "This is not the primary node; do nothing, exiting"
    exit 0
fi
while [ ! -f /usr/lib/hue/desktop/core/src/desktop/management/commands/
desktop_document_cleanup.py ]
```

```
do
  sleep 1
done
sudo systemctl stop hue.service
sudo sed -i 's+ LOG.warn+ # LOG.warn+g' /usr/lib/hue/desktop/core/src/desktop/
management/commands/desktop_document_cleanup.py
sudo /usr/lib/hue/build/env/bin/hue desktop_document_cleanup --keep-days $1
sudo systemctl start hue.service
```

Incompatibilité entre les versions de Hue

Si vous utilisez la même base de données de métadonnées Hue sur plusieurs clusters compatibles Hue, nous recommandons que ces clusters exécutent la même version de Hue. Les différentes versions de Hue peuvent avoir des schémas différents pour la base de données de métadonnées Hue. L'utilisation de la même base de données pour différentes versions peut entraîner l'échec d'une installation de Hue. Par exemple, essayer d'utiliser la même base de données pour deux clusters sur lesquels les versions 4.10.0 et 4.11.0 sont installées peut provoquer des erreurs de connexion pour les utilisateurs qui tentent de se connecter au cluster 4.10.0.

Pour obtenir la liste des labels Amazon et des EMR versions installées correspondantes de Hue, consultez l'[historique des versions de Hue](#).

Connexion à l'interface utilisateur Web de Hue

La connexion à l'interface utilisateur Web Hue est identique à la connexion à n'importe quelle HTTP interface hébergée sur le nœud principal d'un cluster. La procédure suivante décrit comment accéder à l'interface utilisateur de Hue. Pour plus d'informations, consultez la section [Afficher les interfaces Web hébergées sur des EMR clusters](#) dans l'Amazon EMR Management Guide.

Pour afficher l'interface utilisateur Web de Hue

1. Suivez ces instructions pour [configurer un SSH tunnel vers le nœud principal à l'aide de la redirection de port dynamique](#) dans le guide EMR de gestion Amazon.
2. Entrez l'adresse suivante dans votre navigateur pour ouvrir l'interface Web de Hue : `http://master public DNS:8888` où *master public dns* est le DNS nom public de votre nœud maître de cluster, par exemple `ec2-11-22-333-44.compute-1.amazonaws.com`.
3. Sur l'écran de connexion à Hue, si vous êtes administrateur et que vous vous connectez pour la première fois, saisissez un nom d'utilisateur et un mot de passe pour créer votre supercompte

Hue, puis sélectionnez Créer un compte. Sinon, tapez votre nom d'utilisateur et votre mot de passe, puis sélectionnez Create account (Créer un compte), ou saisissez les informations d'identification fournies par votre administrateur.

Utilisation de Hue avec une base de données distante sur Amazon RDS

Par défaut, les informations utilisateur et l'historique des requêtes de Hue sont stockés dans une SQL base de données My locale sur le nœud principal. Vous pouvez également créer un ou plusieurs clusters compatibles HUE à l'aide d'une configuration stockée dans Amazon S3 et d'une base de données My database dans Amazon Relational Database Service (Amazon). RDS Cela vous permet de conserver les informations utilisateur et l'historique des requêtes créés par Hue sans faire fonctionner votre EMR cluster Amazon. Nous vous recommandons d'utiliser le chiffrement côté serveur Amazon S3 pour stocker le fichier de configuration.


Commencez par créer la base de données distante pour Hue.

Pour créer la base de données externe Ma SQL base de données

1. Ouvrez la RDS console Amazon à l'adresse <https://console.aws.amazon.com/rds/>.
2. Cliquez sur Launch a DB Instance (Lancement d'une instance DB).
3. Choisissez Mon, SQL puis cliquez sur Sélectionner.
4. Conservez la sélection par défaut de déploiement multi-AZ et de IOPS stockage provisionné et cliquez sur Suivant.
5. Conservez les valeurs par défaut des spécifications de l'instance, spécifiez les paramètres, puis cliquez sur Next (Suivant).
6. Dans la page Configuration de paramètres avancés, choisissez un groupe de sécurité et un nom de base de données adéquats. Le groupe de sécurité que vous utilisez doit au moins autoriser l'TCPaccès d'entrée au port 3306 depuis le nœud principal de votre cluster. Si vous n'avez pas créé votre cluster à ce stade, vous pouvez autoriser tous les hôtes à se connecter au port 3306 et ajuster le groupe de sécurité, une fois que vous avez lancé le cluster. Cliquez sur Launch DB Instance (Lancement d'une instance DB).
7. Dans le RDS tableau de bord, sélectionnez Instances et sélectionnez l'instance que vous venez de créer. Lorsque votre base de données est disponible, notez le nom de base de données, le nom d'utilisateur, le mot de passe et le nom d'hôte de l'RDSinstance. Vous allez utiliser ces informations pour créer et configurer votre cluster.


Pour spécifier une SQL base de données My externe pour Hue lors du lancement d'un cluster à l'aide du AWS CLI

Pour spécifier une base de SQL données My Database externe pour Hue lors du lancement d'un cluster à l'aide du AWS CLI, utilisez les informations que vous avez indiquées lors de la création de votre RDS instance pour la configurer hue .ini avec un objet de configuration

 Note

Vous pouvez créer plusieurs clusters qui utilisent la même base de données externe, mais chaque cluster partage l'historique des requêtes et les informations utilisateur.

- À l'aide de AWS CLI, créez un cluster sur lequel Hue est installé, en utilisant la base de données externe que vous avez créée et en référant un fichier de configuration avec une classification de configuration pour Hue qui spécifie les propriétés de la base de données. L'exemple suivant crée un cluster avec Hue installé en référant un fichier de configuration dans Amazon S3, `myConfig.json`, qui spécifie la configuration de la base de données.

 Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Hue Name=Spark
Name=Hive \
--instance-type m5.xlarge --instance-count 3 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-
default-roles
```

Vous trouverez ci-dessous des exemples de contenus du fichier `myConfig.json`. Remplacez *dbname*, *username*, *password*, et *RDS instance hostname* avec les valeurs que vous avez indiquées précédemment dans le RDS tableau de bord.

```
[{
  "Classification": "hue-ini",
  "Properties": {},
```

```
"Configurations": [  
  {  
    "Classification": "desktop",  
    "Properties": {},  
    "Configurations": [  
      {  
        "Classification": "database",  
        "Properties": {  
          "name": "dbname",  
          "user": "username",  
          "password": "password",  
          "host": "RDS instance hostname",  
          "port": "3306",  
          "engine": "mysql"  
        },  
        "Configurations": []  
      },  
    ]  
  }  
]
```

Résolution des problèmes

En cas de RDS défaillance d'Amazon

Il est possible que les utilisateurs rencontrent des retards en exécutant une requête parce que l'instance de base de données Hue ne répond pas ou est en cours de basculement. Voici quelques faits et directives pour ce problème :

- Si vous vous connectez à la RDS console Amazon, vous pouvez rechercher des événements de basculement. Par exemple, pour voir si un basculement est en cours ou a eu lieu, recherchez des événements tels que « Le basculement de l'instance Multi-AZ a commencé » et « Basculement de l'instance Multi-AZ terminé. »
- Il faut environ 30 secondes à une RDS instance pour effectuer un basculement.
- Si vous recevez des longer-than-normal réponses à des requêtes dans Hue, essayez de réexécuter la requête.

Configurations avancées pour Hue

Cette section comprend les rubriques suivantes.

Rubriques

- [Configurer Hue pour les LDAP utilisateurs](#)

Configurer Hue pour les LDAP utilisateurs

L'intégration LDAP permet aux utilisateurs de se connecter à Hue en utilisant les informations d'identification existantes stockées dans un LDAP répertoire. Lorsque vous intégrez Hue à HueLDAP, vous n'avez pas besoin de gérer les informations utilisateur de manière indépendante dans Hue. Les informations ci-dessous illustrent l'intégration de Hue à Microsoft Active Directory, mais les options de configuration sont analogues à celles de n'importe quel LDAP annuaire.

LDAP l'authentification se lie d'abord au serveur et établit la connexion. Ensuite, la connexion établie est utilisée pour toute requête ultérieure visant à rechercher des informations sur LDAP l'utilisateur. A moins que votre serveur Active Directory autorise les connexions anonymes, une connexion doit être établie à l'aide d'un nom unique et d'un mot de passe de liaison. Le nom unique (ou DN) de liaison est défini par le paramètre de configuration `bind_dn`. Le mot de passe de liaison est défini par le paramètre de configuration `bind_password`. Hue propose deux méthodes pour lier les LDAP requêtes : liaison de recherche et liaison directe. La méthode préférée pour utiliser Hue avec Amazon EMR est le `search bind`.

Lorsque la liaison de recherche est utilisée avec Active Directory, Hue utilise l'attribut de nom d'utilisateur (défini par `user_name_attr` config) pour rechercher l'attribut qui doit être extrait du nom unique (ou DN) de la base. La liaison de recherche est utile lorsque le nom unique complet n'est pas connu pour l'utilisateur de Hue.

Par exemple, vous avez peut-être `user_name_attr` config définir pour utiliser le nom commun (ou CN). Dans ce cas, le serveur Active Directory utilise le nom d'utilisateur Hue fourni au cours de l'ouverture de session pour rechercher dans l'arborescence de répertoires un nom commun qui corresponde, en commençant par le nom unique de la base. Si le nom commun pour l'utilisateur Hue est trouvé, le nom unique de l'utilisateur est renvoyé par le serveur. Hue construit ensuite un nom unique utilisé pour authentifier l'utilisateur en effectuant une opération de liaison.

Note

La liaison de recherche parcourt toutes les sous-arborescences du répertoire pour trouver des noms d'utilisateur, en commençant par le nom unique de la base. Le nom distinctif de base spécifié dans la LDAP configuration Hue doit être le parent le plus proche du nom d'utilisateur, sinon vos performances LDAP d'authentification risquent d'en pâtir.

Lorsque la liaison directe est utilisée avec Active Directory, le `nt_domain` ou `ldap_username_pattern` exact doit être utilisé pour s'authentifier. Lorsque la liaison directe est utilisée, si l'attribut de domaine `nt` (défini par le paramètre de configuration `nt_domain`) est défini, un modèle de nom unique d'utilisateur est créé à l'aide du formulaire : `<login username>@nt_domain`. Ce modèle est utilisé pour rechercher dans toutes les sous-arborescences du répertoire, en commençant par le nom unique de la base. Si le domaine `nt` n'est pas configuré, Hue recherche un modèle exact de nom unique pour l'utilisateur (défini par le paramètre de configuration `ldap_username_pattern`). Dans ce cas, le serveur recherche une valeur `ldap_username_pattern` correspondante dans toutes les sous-arborescences du répertoire, en commençant par le nom unique de la base.

Pour lancer un cluster doté de LDAP propriétés pour Hue à l'aide du AWS CLI

- Pour spécifier les LDAP propriétés `dehue-ini`, créez un cluster sur lequel Hue est installé et référez un fichier json contenant les propriétés de configuration pourLDAP. Un exemple de commande est présenté ci-dessous. Elle fait référence à un fichier de configuration `myConfig.json` stocké dans Amazon S3.

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Hue Name=Spark
Name=Hive \
--instance-type m5.xlarge --instance-count 3 --configurations https://
s3.amazonaws.com/mybucket/myfolder/myConfig.json.
```

Vous trouverez ci-dessous des exemples de contenus `myConfig.json`.

```
[
  {
    "Classification": "hue-ini",
    "Properties": {},
    "Configurations": [
      {
```

```

    "Classification": "desktop",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "ldap",
        "Properties": {},
        "Configurations": [
          {
            "Classification": "ldap_servers",
            "Properties": {},
            "Configurations": [
              {
                "Classification": "yourcompany",
                "Properties": {
                  "base_dn":
"DC=yourcompany,DC=hue,DC=com",
                  "ldap_url": "ldap://ldapurl",
                  "search_bind_authentication": "true",
                  "bind_dn":
"CN=hue,CN=users,DC=yourcompany,DC=hue,DC=com",
                  "bind_password": "password"
                },
                "Configurations": []
              }
            ]
          }
        ]
      },
      {
        "Classification": "auth",
        "Properties": {
          "backend": "desktop.auth.backend.LdapBackend"
        }
      }
    ]
  }
]

```

Note

Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier des classifications de configuration supplémentaires pour chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, utilisez la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).

Pour afficher LDAP les paramètres dans Hue

1. Vérifiez que vous disposez d'une VPN connexion ou d'un SSH tunnel actif avec le nœud principal du EMR cluster Amazon. Ensuite, dans votre navigateur, tapez *master-public-dns:8888* pour ouvrir l'interface Web de Hue.
2. Connectez-vous à l'aide de vos informations d'identification d'administrateur Hue. Si la fenêtre Did you know? (Le saviez vous ?) s'ouvre, cliquez sur Got it, prof! (Compris, professeur !) pour la fermer.
3. Cliquez sur l'icône Hue dans la barre d'outils.
4. Sur la page About Hue (À propos de Hue), sélectionnez Configuration.
5. Dans la section Configuration Sections and Variables (Sections et variables de configuration), cliquez sur Desktop (Bureau).
6. Défilez à la section ldap pour afficher vos paramètres.

Historique des versions de Hue

Le tableau suivant répertorie la version de Hue incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Hue

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-7.2.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.36.2	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-7.1.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server,

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
		mariadb-server, oozie-client, oozie-server
emr-7.0.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.15.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.14.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.13.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.12.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.11.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.11.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.10.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.10.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.9.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.9.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.8.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.8.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.7.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.36.1	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.36.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.6.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.35.0	4.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.5.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.4.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.3.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.3.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.2.1	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.2.0	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.1.1	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.1.0	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-6.0.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-6.0.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.34.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.33.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.33.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.32.1	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.32.0	4.8.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.31.1	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.31.0	4.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.30.2	4.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.30.1	4.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server
emr-5.30.0	4.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
mr-5.29.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.28.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.28.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.27.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.27.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.26.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.25.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.24.1	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.24.0	4.4.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.23.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.23.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.22.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.21.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.21.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.21.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.20.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.20.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.19.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.19.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.18.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.18.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.17.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.17.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.17.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.16.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.16.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.15.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.15.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.14.2	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.14.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.14.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.13.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.13.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.3	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.12.2	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.12.1	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.12.0	4.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.4	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.11.3	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.2	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.11.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.11.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.10.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.10.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.9.1	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.9.0	4.0.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.8.3	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.2	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.8.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.8.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.7.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.7.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.6.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.6.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.4	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.5.3	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.2	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.5.1	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.5.0	3.12.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.4.1	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.4.0	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.3.2	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.3.1	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.3.0	3.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.2.3	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.2.2	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.2.1	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.2.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.1.1	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-5.1.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.0.3	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-5.0.2	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-5.0.1	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-5.0.0	3.10.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.9.6	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.5	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-4.9.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.9.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-4.9.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.5	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-4.8.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.8.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-4.8.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.7.4	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.7.3	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-4.7.2	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-client, oozie-server
emr-4.7.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.7.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-4.6.1	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.6.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.5.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-4.4.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.3.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server
emr-4.2.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Étiquette Amazon EMR Release	Version Hue	Composants installés avec Hue
emr-4.1.0	3.7.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hue-server, mysql-server, oozie-server

Iceberg

[Apache Iceberg](#) est un format de tableau ouvert pour les grands jeux de données dans Amazon Simple Storage Service (Amazon S3). Il fournit des performances de requête rapides sur de grandes tables, des validations atomiques, des écritures SQL simultanées et une évolution de table compatible. À partir d'Amazon EMR 6.5.0, vous pouvez utiliser Apache Spark 3 sur des EMR clusters Amazon au format de table Iceberg.

Le tableau suivant répertorie la version d'Iceberg incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Iceberg.

Pour la version des composants installés avec Iceberg dans cette version, voir [Versions des composants de la version 7.2.0](#).

Informations sur la version d'Iceberg pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Iceberg	Composants installés avec Iceberg
emr-7.2.0	Iceberg 1.5.0-amzn-0	Not available.

Le tableau suivant répertorie la version d'Iceberg incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Iceberg.

Pour connaître la version des composants installés avec Iceberg dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version d'Iceberg pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Iceberg	Composants installés avec Iceberg
emr-6.15.0	Iceberg 1.4.0-amzn-0	Not available.

Rubriques

- [Comment fonctionne Iceberg](#)
- [Utiliser un cluster sur lequel Iceberg est installé](#)

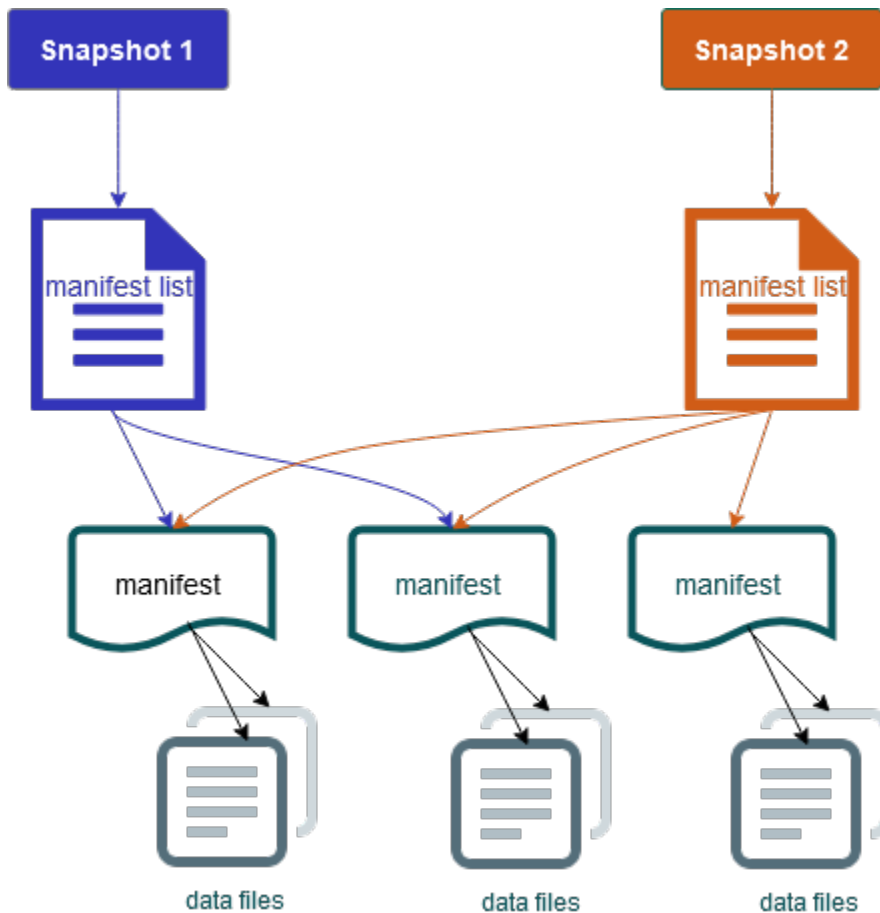
- [Considérations et limites relatives à l'utilisation d'Iceberg sur Amazon EMR](#)
- [Historique des versions d'Iceberg](#)

Comment fonctionne Iceberg

Iceberg suit les fichiers de données individuels dans une table plutôt que dans des répertoires. Ainsi, les rédacteurs peuvent créer des fichiers de données sur place (les fichiers ne sont ni déplacés ni modifiés). De plus, les rédacteurs ne peuvent ajouter des fichiers à la table que dans le cadre d'un commit explicite. L'état de la table est conservé dans les fichiers de métadonnées. Toutes les modifications apportées à l'état de la table créent un nouveau fichier de métadonnées qui remplace automatiquement les anciennes métadonnées. Le fichier de métadonnées de table suit le schéma de la table, la configuration du partitionnement et d'autres propriétés.

Il inclut également des instantanés du contenu du tableau. Chaque instantané est un ensemble complet de fichiers de données figurant dans le tableau à un moment donné. Les instantanés sont répertoriés dans le fichier de métadonnées, mais les fichiers d'un instantané sont stockés dans des fichiers manifestes distincts. Les transitions atomiques d'un fichier de métadonnées de table à l'autre permettent d'isoler les instantanés. Les lecteurs utilisent l'instantané actuel lorsqu'ils ont chargé les métadonnées de la table. Les lecteurs ne sont pas affectés par les modifications tant qu'ils n'ont pas actualisé et sélectionné un nouvel emplacement de métadonnées. Les fichiers de données des instantanés sont stockés dans un ou plusieurs fichiers manifestes qui contiennent une ligne pour chaque fichier de données de la table, ses données de partition et ses métriques. Un instantané est l'union de tous les fichiers contenus dans ses manifestes. Les fichiers manifestes peuvent également être partagés entre les instantanés afin d'éviter de réécrire des métadonnées peu modifiées.

Schéma d'instantané d'Iceberg



Iceberg offre les fonctionnalités suivantes :

- Prend en charge ACID les transactions et les voyages dans le temps dans votre lac de données Amazon S3.
- Les nouvelles tentatives de validation bénéficient des avantages en termes de performances d'une [simultanéité optimiste](#).
- La résolution des conflits au niveau des fichiers entraîne une forte simultanéité.
- Avec des statistiques min-max par colonne dans les métadonnées, vous pouvez ignorer des fichiers, ce qui améliore les performances des requêtes sélectives.
- Vous pouvez organiser les tables selon des configurations de partition flexibles, l'évolution des partitions permettant de mettre à jour les schémas de partition. Les requêtes et les volumes de données peuvent alors changer sans recourir à des répertoires physiques.
- Prend en charge l'[évolution des schémas](#) et leur application.
- Les tables Iceberg agissent comme des puits idempotents et des sources rejouables. Cela permet le streaming et le support par lots avec des pipelines uniques. Les récepteurs idempotents suivent

les opérations d'écriture réussies dans le passé. Par conséquent, le récepteur peut à nouveau demander des données en cas de panne et supprimer des données si elles ont été envoyées plusieurs fois.

- Consultez l'historique et le lignage, y compris l'évolution des tables, l'historique des opérations et les statistiques pour chaque validation.
- Migrez à partir d'un ensemble de données existant avec un choix de format de données (ParquetORC, Avro) et de moteur d'analyse (Spark, Trino, PrestoDB, Flink, Hive).

Utiliser un cluster sur lequel Iceberg est installé

Cette section contient des informations sur l'utilisation d'Iceberg avec Spark, Trino, Flink et Hive.

Utiliser un cluster Iceberg avec Spark

À partir de EMR la version 6.5.0 d'Amazon, vous pouvez utiliser Iceberg avec votre cluster Spark sans avoir à inclure d'actions bootstrap. Pour EMR les versions 6.4.0 et antérieures d'Amazon, vous pouvez utiliser une action bootstrap pour préinstaller toutes les dépendances nécessaires.

Dans ce didacticiel, vous allez utiliser le AWS CLI pour travailler avec Iceberg sur un cluster Amazon EMR Spark. Pour utiliser la console afin de créer un cluster sur lequel Iceberg est installé, suivez les étapes décrites dans [Créer un lac de données Apache Iceberg à l'aide d'Amazon Athena, EMR AWS Amazon et Glue](#).

Création d'un cluster Iceberg

Vous pouvez créer un cluster avec Iceberg installé à l'aide du AWS Management Console, du AWS CLI ou d'Amazon EMRAPI. Dans ce didacticiel, vous allez utiliser le AWS CLI pour travailler avec Iceberg sur un EMR cluster Amazon. Pour utiliser la console afin de créer un cluster sur lequel Iceberg est installé, suivez les étapes décrites dans [Créer un lac de données Apache Iceberg à l'aide d'Amazon Athena, EMR AWS Amazon et Glue](#).

Pour utiliser Iceberg sur Amazon EMR avec le AWS CLI, créez d'abord un cluster en suivant les étapes suivantes. Pour plus d'informations sur la spécification de la classification des icebergs à l'aide du AWS CLI, voir [Fournissez une configuration à l'aide du AWS CLI lorsque vous créez un cluster](#) ou [Fournissez une configuration à l'aide de Java SDK lorsque vous créez un cluster](#).

1. Créez un fichier `configurations.json` avec le contenu suivant :


```
[{
  "Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

2. Créez ensuite un cluster à l'aide de la configuration suivante. Remplacez l'exemple de chemin de compartiment Amazon S3 et l'ID de sous-réseau par les vôtres.

```
aws emr create-cluster --release-label emr-6.5.0 \
--applications Name=Spark \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_Spark_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole_V2 \
--ec2-attributes
  InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef0
```

Vous pouvez également créer un EMR cluster Amazon incluant l'application Spark et inclure le fichier en `/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` tant que JAR dépendance dans une tâche Spark. Pour plus d'informations, consultez [Soumission d'applications](#).

Pour inclure le fichier jar en tant que dépendance dans une tâche Spark, ajoutez la propriété de configuration suivante à l'application Spark :

```
--conf "spark.jars=/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar"
```

Pour plus d'informations sur les dépendances des tâches Spark, consultez la section [Gestion des dépendances](#) dans le document d'Apache Spark [Running Spark on Kubernetes](#).

Initialisation d'une session Spark pour Iceberg

Les exemples suivants montrent comment lancer le shell interactif Spark, utiliser Spark submit ou utiliser Amazon EMR Notebooks pour travailler avec Iceberg sur Amazon. EMR

spark-shell

1. Connectez-vous au nœud principal à l'aide deSSH. Pour plus d'informations, consultez [Connect to the master node SSH à l'aide](#) de l'Amazon EMR Management Guide.
2. Entrez la commande suivante pour lancer le shell Spark. Pour utiliser la PySpark coque, remplacez-la `spark-shell` par `pySpark`.

```
spark-shell \  
--conf  
  "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions"  
 \  
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \  
--conf "spark.sql.catalog.dev.type=hadoop" \  
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

spark-submit

1. Connectez-vous au nœud principal à l'aide deSSH. Pour plus d'informations, consultez [Connect to the master node SSH à l'aide](#) de l'Amazon EMR Management Guide.
2. Entrez la commande suivante pour lancer la session Spark pour Iceberg.

```
spark-submit \  
--conf  
  "spark.sql.extensions=org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions"  
 \  
--conf "spark.sql.catalog.dev=org.apache.iceberg.spark.SparkCatalog" \  
--conf "spark.sql.catalog.dev.type=hadoop" \  
--conf "spark.sql.catalog.dev.warehouse=s3://DOC-EXAMPLE-BUCKET/example-prefix/"
```

EMR Studio notebooks

Pour initialiser une session Spark à l'aide de blocs-notes EMR Studio, configurez votre session Spark à l'aide de la commande `%%configure` magique de votre EMR bloc-notes Amazon, comme dans l'exemple suivant. Pour plus d'informations, consultez la section [Utiliser la magie EMR des blocs-notes dans](#) le guide de EMRgestion Amazon.

```
%%configure -f  
{  
  "conf":{
```

```

"spark.sql.extensions":"org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions",
"spark.sql.catalog.dev":"org.apache.iceberg.spark.SparkCatalog",
"spark.sql.catalog.dev.type":"hadoop",
"spark.sql.catalog.dev.warehouse":"s3://DOC-EXAMPLE-BUCKET/example-prefix/"
}
}

```

Écriture dans une table Iceberg

L'exemple suivant montre comment créer un jeu de données Iceberg DataFrame et l'écrire sous forme de jeu de données. Les exemples montrent comment travailler avec des ensembles de données à l'aide du shell Spark lorsque vous êtes connecté au nœud principal en SSH tant qu'utilisateur hadoop par défaut.

Note

Pour coller des exemples de code dans le shell Spark, tapez `:paste` à l'invite, collez l'exemple, puis appuyez sur CTRL+D.

PySpark

Spark inclut un shell basé sur Python, `pyspark`, que vous pouvez utiliser pour créer un prototype de programmes Spark écrits en Python. Invoquez `pyspark` sur le nœud principal.

```

## Create a DataFrame.
data = spark.createDataFrame([
    ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
    ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
    ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
    ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")
],[ "id", "creation_date", "last_update_time"])

## Write a DataFrame as a Iceberg dataset to the Amazon S3 location.
spark.sql("""CREATE TABLE IF NOT EXISTS dev.db.iceberg_table (id string,
creation_date string,
last_update_time string)
USING iceberg
location 's3://DOC-EXAMPLE-BUCKET/example-prefix/db/iceberg_table'""")

```

```
data.writeTo("dev.db.iceberg_table").append()
```

Scala

```
import org.apache.spark.sql.SaveMode
import org.apache.spark.sql.functions._

// Create a DataFrame.
val data = Seq(
  ("100", "2015-01-01", "2015-01-01T13:51:39.340396Z"),
  ("101", "2015-01-01", "2015-01-01T12:14:58.597216Z"),
  ("102", "2015-01-01", "2015-01-01T13:51:40.417052Z"),
  ("103", "2015-01-01", "2015-01-01T13:51:40.519832Z")
).toDF("id", "creation_date", "last_update_time")

// Write a DataFrame as a Iceberg dataset to the Amazon S3 location.
spark.sql("""CREATE TABLE IF NOT EXISTS dev.db.iceberg_table (id string,
creation_date string,
last_update_time string)
USING iceberg
location 's3://DOC-EXAMPLE-BUCKET/example-prefix/db/iceberg_table'""")

data.writeTo("dev.db.iceberg_table").append()
```

Lecture à partir d'une table Iceberg

PySpark

```
df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

Scala

```
val df = spark.read.format("iceberg").load("dev.db.iceberg_table")
df.show()
```

Spark SQL

```
SELECT * from dev.db.iceberg_table LIMIT 10
```

Configurer les propriétés de Spark pour utiliser le catalogue de données AWS Glue comme métastore de tables Iceberg

Pour utiliser le catalogue AWS Glue comme métastore pour les tables Iceberg, définissez les propriétés de configuration de Spark comme suit :

```
spark-submit \  
  --conf spark.sql.catalog.my_catalog=org.apache.iceberg.spark.SparkCatalog \  
  --conf spark.sql.catalog.my_catalog.warehouse=s3://<bucket>/<prefix> \  
  --conf spark.sql.catalog.my_catalog.catalog-  
impl=org.apache.iceberg.aws.glue.GlueCatalog \  
  --conf spark.sql.catalog.my_catalog.io-impl=org.apache.iceberg.aws.s3.S3FileIO \  
  --conf spark.sql.catalog.my_catalog.lock-  
impl=org.apache.iceberg.aws.dynamodb.DynamoDbLockManager \  
  --conf spark.sql.catalog.my_catalog.lock.table=myGlueLockTable
```

Utiliser un cluster Iceberg avec Trino

À partir de EMR la version 6.6.0 d'Amazon, vous pouvez utiliser Iceberg avec votre cluster Trino.

Dans ce didacticiel, vous allez utiliser le AWS CLI pour travailler avec Iceberg sur un cluster Amazon EMR Trino. Pour utiliser la console afin de créer un cluster sur lequel Iceberg est installé, suivez les étapes décrites dans [Créer un lac de données Apache Iceberg à l'aide d'Amazon Athena, EMR AWS Amazon et Glue](#).

Création d'un cluster Iceberg

Pour utiliser Iceberg sur Amazon EMR avec le AWS CLI, créez d'abord un cluster en suivant les étapes suivantes. Pour plus d'informations sur la spécification de la classification des icebergs à l'aide du AWS CLI, voir [Fournissez une configuration à l'aide du AWS CLI lorsque vous créez un cluster](#) ou [Fournissez une configuration à l'aide de Java SDK lorsque vous créez un cluster](#).

1. Créez un fichier `iceberg.properties` et définissez une valeur pour le catalogue de votre choix. Par exemple, si vous souhaitez utiliser le métastore Hive comme catalogue, le contenu de votre fichier doit être le suivant.

```
connector.name=iceberg  
hive.metastore.uri=thrift://localhost:9083
```

Si vous souhaitez utiliser le AWS Glue Data Catalog comme boutique, le contenu de votre fichier doit être le suivant.

```
connector.name=iceberg
iceberg.catalog.type=glue
```

2. Créez une action d'amorçage qui copie iceberg.properties depuis Amazon S3 vers /etc/trino/conf/catalog/iceberg.properties, comme dans l'exemple suivant. Pour plus d'informations sur les actions d'amorçage, consultez la section [Création d'actions d'amorçage pour installer des logiciels supplémentaires](#).

```
set -ex
sudo aws s3 cp s3://DOC-EXAMPLE-BUCKET/iceberg.properties /etc/trino/conf/catalog/iceberg.properties
```

3. Créez un cluster avec la configuration suivante, en remplaçant l'exemple de chemin du script d'actions amorçage et le nom de clé par les vôtres.

```
aws emr create-cluster --release-label emr-6.7.0 \
--applications Name=Trino \
--region us-east-1 \
--name My_Trino_Iceberg_Cluster \
--bootstrap-actions '[{"Path":"s3://DOC-EXAMPLE-BUCKET", "Name":"Add
iceberg.properties"}]' \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=c3.4xlarge
InstanceGroupType=CORE,InstanceCount=3,InstanceType=c3.4xlarge \
--use-default-roles \
--ec2-attributes KeyName=<key-name>
```

Initialisation d'une session Trino pour Iceberg

Pour initialiser une session Trino, exécutez la commande suivante.

```
trino-cli --catalog iceberg
```

Écriture dans une table Iceberg

Créez et écrivez dans votre table à l'aide des SQL commandes suivantes.

```
trino> SHOW SCHEMAS;
trino> CREATE TABLE default.iceberg_table (
    id int,
    data varchar,
    category varchar)
WITH (
    format = 'PARQUET',
    partitioning = ARRAY['category', 'bucket(id, 16)'],
    location = 's3://DOC-EXAMPLE-BUCKET/<prefix>')

trino> INSERT INTO default.iceberg_table VALUES (1,'a','c1'), (2,'b','c2'),
(3,'c','c3');
```

Lire à partir d'une table pour Iceberg

Pour lire les données de votre table Iceberg, exécutez la commande suivante.

```
trino> SELECT * from default.iceberg_table;
```

Utiliser un cluster Iceberg avec Flink

À partir de EMR la version 6.9.0 d'Amazon, vous pouvez utiliser Iceberg avec un cluster Flink sans avoir à effectuer les étapes de configuration requises lors de l'utilisation de l'intégration open source Iceberg Flink.

Création d'un cluster Iceberg

Vous pouvez créer un cluster avec Iceberg installé à l'aide du AWS Management Console AWS CLI, du ou d'Amazon EMRAPI. Dans ce didacticiel, vous allez utiliser le AWS CLI pour travailler avec Iceberg sur un EMR cluster Amazon. Pour utiliser la console afin de créer un cluster sur lequel Iceberg est installé, suivez les étapes décrites dans [Créer un lac de données Apache Iceberg à l'aide d'Amazon Athena, EMR AWS Amazon et Glue](#).

Pour utiliser Iceberg sur Amazon EMR avec le AWS CLI, créez d'abord un cluster en suivant les étapes suivantes. Pour plus d'informations sur la spécification de la classification des icebergs à l'aide du AWS CLI, voir [Fournissez une configuration à l'aide du AWS CLI lorsque vous créez un cluster](#) ou [Fournissez une configuration à l'aide de Java SDK lorsque vous créez un cluster](#). Créez un fichier nommé `configurations.json` avec le contenu suivant :

```
[{
```

```
"Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

Créez ensuite un cluster avec la configuration suivante, en remplaçant les exemples de chemin de compartiment et d'ID de sous-réseau Amazon S3 par vos propres valeurs :

```
aws emr create-cluster --release-label emr-6.9.0 \
--applications Name=Flink \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_flink_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef
```

Vous pouvez également créer un cluster Amazon EMR 6.9.0 contenant une application Flink et utiliser le fichier `/usr/share/aws/iceberg/lib/iceberg-flink-runtime.jar` comme JAR dépendance dans une tâche Flink.

Utilisation du client Flink SQL

Le script SQL client se trouve sous `/usr/lib/flink/bin`. Vous pouvez exécuter le script à l'aide de la commande suivante :

```
flink-yarn-session -d # starting the Flink YARN Session in detached mode
./sql-client.sh
```

Cela lance un Flink SQL Shell.

Exemples de Flink

Création d'une table Iceberg

Flink SQL

```
CREATE CATALOG glue_catalog WITH (
  'type'='iceberg',
  'warehouse'='<WAREHOUSE>',
  'catalog-impl'='org.apache.iceberg.aws.glue.GlueCatalog',
```



```

    'io-impl'='org.apache.iceberg.aws.s3.S3FileIO',
    'lock-impl'='org.apache.iceberg.aws.dynamodb.DynamoDbLockManager',
    'lock.table'='myGlueLockTable'
);

USE CATALOG glue_catalog;

CREATE DATABASE IF NOT EXISTS <DB>;

USE <DB>;

CREATE TABLE IF NOT EXISTS `glue_catalog`.`<DB>`.`sample` (id int, data string);

```

Tableau API

```

EnvironmentSettings settings =
    EnvironmentSettings.newInstance().inBatchMode().build();

TableEnvironment tEnv = TableEnvironment.create(settings);

String warehouse = "<WAREHOUSE>";
String db = "<DB>";

tEnv.executeSql(
    "CREATE CATALOG glue_catalog WITH (\n"
        + "    'type'='iceberg',\n"
        + "    'warehouse'='"
        + warehouse
        + "',\n"
        + "    'catalog-impl'='org.apache.iceberg.aws.glue.GlueCatalog',
\n"
        + "    'io-impl'='org.apache.iceberg.aws.s3.S3FileIO'\n"
        + " );");

tEnv.executeSql("USE CATALOG glue_catalog;");
tEnv.executeSql("CREATE DATABASE IF NOT EXISTS " + db + ";");
tEnv.executeSql("USE " + db + ";");
tEnv.executeSql(
    "CREATE TABLE `glue_catalog`.`" + db + "`.`sample` (id bigint, data string);");

```

Écriture dans une table Iceberg

Flink SQL

```
INSERT INTO `glue_catalog`.`<DB>`.`sample` values (1, 'a'),(2,'b'),(3,'c');
```

Tableau API

```
tEnv.executeSql(  
    "INSERT INTO `glue_catalog`.`"  
    + db  
    + "`.`sample` values (1, 'a'),(2,'b'),(3,'c');");
```

Flux de données API

```
final StreamExecutionEnvironment env =  
    StreamExecutionEnvironment.getExecutionEnvironment();  
  
final StreamTableEnvironment tableEnv = StreamTableEnvironment.create(env);  
  
String db = "<DB Name>";  
  
String warehouse = "<Warehouse Path>";  
  
GenericRowData rowData1 = new GenericRowData(2);  
rowData1.setField(0, 1L);  
rowData1.setField(1, StringData.fromString("a"));  
  
DataStream<RowData> input = env.fromElements(rowData1);  
  
Map<String, String> props = new HashMap<>();  
props.put("type", "iceberg");  
props.put("warehouse", warehouse);  
props.put("io-impl", "org.apache.iceberg.aws.s3.S3FileIO");  
  
CatalogLoader glueCatlogLoader =  
    CatalogLoader.custom(  
        "glue",  
        props,  
        new Configuration(),  
        "org.apache.iceberg.aws.glue.GlueCatalog");  
  
TableLoader tableLoader =  
    TableLoader.fromCatalog(glueCatlogLoader, TableIdentifier.of(db, "sample"));  
  
DataStreamSink<Void> dataStreamSink =
```

```
FlinkSink.forRowData(input).tableLoader(tableLoader).append();

env.execute("Datastream Write");
```

Lecture à partir d'une table Iceberg

Flink SQL

```
SELECT * FROM `glue_catalog`.`<DB>`.`sample`;
```

Tableau API

```
Table result = tEnv.sqlQuery("select * from `glue_catalog`.`" + db + "`.`sample`");
```

Flux de données API

```
final StreamExecutionEnvironment env =
    StreamExecutionEnvironment.getExecutionEnvironment();

final StreamTableEnvironment tableEnv = StreamTableEnvironment.create(env);

String db = "<DB Name>";

String warehouse = "<Warehouse Path>";

Map<String, String> props = new HashMap<>();
props.put("type", "iceberg");
props.put("warehouse", warehouse);
props.put("io-impl", "org.apache.iceberg.aws.s3.S3FileIO");

CatalogLoader glueCatalogLoader =
    CatalogLoader.custom(
        "glue",
        props,
        new Configuration(),
        "org.apache.iceberg.aws.glue.GlueCatalog");

TableLoader tableLoader =
    TableLoader.fromCatalog(glueCatalogLoader, TableIdentifier.of(db, "sample"));

DataStream<RowData> batch =

    FlinkSource.forRowData().env(env).tableLoader(tableLoader).streaming(false).build();
```

```
batch.print().name("print-sink");
```

Utilisation du catalogue Hive

Assurez-vous que les dépendances de Flink et Hive sont résolues comme décrit dans [Configuration de Flink avec Hive Metastore et Glue Catalog](#).

Exécution d'une tâche Flink

L'une des manières de soumettre une tâche à Flink consiste à utiliser une session Flink YARN par tâche. Vous pouvez le lancer à l'aide de la commande suivante :

```
sudo flink run -m yarn-cluster -p 4 -yjm 1024m -ytm 4096m $JAR_FILE_NAME
```

Utiliser un cluster Iceberg avec Hive

Avec les EMR versions 6.9.0 et supérieures d'Amazon, vous pouvez utiliser Iceberg avec un cluster Hive sans avoir à effectuer les étapes de configuration requises pour l'intégration Open Source d'Iceberg Hive. Pour les EMR versions 6.8.0 et antérieures d'Amazon, vous pouvez utiliser une action bootstrap pour installer `iceberg-hive-runtime` JAR afin de configurer le support de Hive for Iceberg.

Amazon EMR 6.9.0 inclut toutes les fonctionnalités nécessaires à l'[intégration de Hive 3.1.3 à Iceberg 0.14.1](#) et inclut également des fonctionnalités EMR ajoutées par Amazon, telles que la sélection automatique des moteurs d'exécution pris en charge lors de l'exécution (Amazon sur 6.9.0). EMR EKS

Création d'un cluster Iceberg

Vous pouvez créer un cluster avec Iceberg installé à l'aide du AWS Management Console, du AWS CLI ou d'Amazon EMRAPI. Dans ce didacticiel, vous allez utiliser le AWS CLI pour travailler avec Iceberg sur un EMR cluster Amazon. Pour utiliser la console afin de créer un cluster sur lequel Iceberg est installé, suivez les étapes décrites dans la section [Créer un lac de données Iceberg à l'aide d'Amazon Athena, EMR AWS Amazon et Glue](#).

Pour utiliser Iceberg sur Amazon EMR avec le AWS CLI, créez d'abord un cluster en suivant les étapes ci-dessous. Pour plus d'informations sur la spécification de la classification Iceberg à l'aide du AWS CLI ou du JavaSDK, voir [Fournissez une configuration à l'aide du AWS CLI lorsque vous créez](#)

[un cluster](#) ou [Fournissez une configuration à l'aide de Java SDK lorsque vous créez un cluster](#). Créez un fichier nommé `configurations.json` avec le contenu suivant:

```
[{
  "Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

Créez ensuite un cluster avec la configuration suivante, en remplaçant l'exemple de chemin de compartiment Amazon S3 et l'ID de sous-réseau par les vôtres :

```
aws emr create-cluster --release-label emr-6.9.0 \
--applications Name=Hive \
--configurations file://iceberg_configurations.json \
--region us-east-1 \
--name My_hive_Iceberg_Cluster \
--log-uri s3://DOC-EXAMPLE-BUCKET/ \
--instance-type m5.xlarge \
--instance-count 2 \
--service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-1234567890abcdef
```

Un cluster Hive Iceberg effectue les opérations suivantes :

- Charge le fichier JAR d'exécution Iceberg Hive dans Hive et active la configuration associée à Iceberg pour le moteur Hive.
- Active la sélection dynamique du moteur d'exécution d'Amazon EMR Hive afin d'empêcher les utilisateurs de configurer le moteur d'exécution pris en charge pour la compatibilité avec Iceberg.

Note

Les clusters Hive Iceberg ne sont actuellement pas compatibles avec AWS Glue Data Catalog. Le catalogue Iceberg par défaut est HiveCatalog, ce qui correspond au métastore configuré pour l'environnement Hive. Pour plus d'informations sur la gestion des catalogues, consultez la section [Utilisation HCatalog](#) dans la [documentation d'Apache Hive](#).

Prise en charge des fonctionnalités

Amazon EMR 6.9.0 prend en charge Hive 3.1.3 et Iceberg 0.14.1. La prise en charge des fonctionnalités est limitée aux fonctionnalités compatibles avec Iceberg pour Hive 3.1.2 et 3.1.3. Les commandes suivantes sont prises en charge :

- Avec les EMR versions 6.9.0 à 6.12.x d'Amazon, vous devez inclure le fichier `libfb303.jar` dans le répertoire Hive. `auxlib` Utilisez la commande suivante pour l'inclure :

```
sudo /usr/bin/ln -sf /usr/lib/hive/lib/libfb303-*.jar /usr/lib/hive/auxlib/libfb303.jar
```

Avec les EMR versions 6.13 et supérieures d'Amazon, le `libfb303` fichier jar est automatiquement lié au répertoire `auxlib` Hive.

- Création d'une table
 - Table non partitionnée – Les tables externes dans Hive peuvent être créées en fournissant le gestionnaire de stockage comme suit :

```
CREATE EXTERNAL TABLE x (i int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

- Table partitionnée – Les tables partitionnées externes dans Hive peuvent être créées comme suit :

```
CREATE EXTERNAL TABLE x (i int) PARTITIONED BY (j int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

Note

Le format de `STORED AS fichierORC/AVRO/n'PARQUET` est pas pris en charge dans Hive 3. La seule option par défaut est Parquet.

- Suppression d'une table – La commande `DROP TABLE` est utilisée pour supprimer des tables, comme dans l'exemple suivant :

```
DROP TABLE [IF EXISTS] table_name [PURGE];
```

- Lecture d'un tableau : Les instructions SELECT peuvent être utilisées pour lire les tables Iceberg dans Hive, comme dans l'exemple suivant. Les moteurs d'exécution pris en charge sont MR et Tez.

```
SELECT * FROM table_name
```

Pour plus d'informations sur la syntaxe de sélection de Hive, voir [LanguageManual Select](#). Pour plus d'informations sur les instructions select associées aux tables Iceberg dans Hive, consultez [Apache Iceberg Select](#).

- Insertion dans une table – L'instruction de HiveQL INSERT INTO fonctionne sur les tables Iceberg avec le support du moteur d'exécution Map Reduce uniquement. EMR Les utilisateurs d'Amazon n'ont pas besoin de définir explicitement le moteur d'exécution, car Amazon EMR Hive sélectionne le moteur pour Iceberg Tables au moment de l'exécution.
- Insert into d'une seul tableau – Exemple :

```
INSERT INTO table_name VALUES ('a', 1);  
INSERT INTO table_name SELECT...
```

- Insert into multitable – Les instructions d'insertion multitable non atomiques dans des instructions sont prises en charge. Exemple :

```
FROM source  
INSERT INTO table_1 SELECT a, b  
INSERT INTO table_2 SELECT c,d;
```

Considérations et limites relatives à l'utilisation d'Iceberg sur Amazon EMR

Cette section inclut les considérations et les limites relatives à l'utilisation d'Iceberg avec Spark, Trino, Flink et Hive.

Considérations relatives à l'utilisation d'Iceberg avec Spark

- Amazon EMR 6.5.0 ne prend pas en charge l'exécution d'Iceberg EMR sur Amazon EKS par défaut. Une image personnalisée Amazon EMR 6.5.0 est disponible afin que vous puissiez la transmettre `--jars local:///usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` en `spark-submit` paramètre pour créer des tables Iceberg sur Amazon EMR

le. EKS Pour plus d'informations, consultez la section [Soumettre une charge de travail Spark dans Amazon à l'EMR](#) aide d'une image personnalisée dans le guide Amazon EMR on EKS Development. Vous pouvez également nous contacter AWS Support pour obtenir de l'aide. À partir d'Amazon EMR 6.6.0, Iceberg est pris en charge sur Amazon EMR le. EKS

- Lorsque vous utilisez AWS Glue comme catalogue pour Iceberg, assurez-vous que la base de données dans laquelle vous créez une table existe dans AWS Glue. Si vous utilisez des services tels que AWS Lake Formation et que vous ne parvenez pas à charger le catalogue, assurez-vous de disposer d'un accès approprié au service pour exécuter la commande.

Considérations relatives à l'utilisation d'Iceberg avec Trino

- Amazon EMR 6.5 ne prend pas en charge le catalogue Trino Iceberg pour Iceberg en mode natif. Trino a besoin d'Iceberg v0.11. Nous vous recommandons donc de lancer un EMR cluster Amazon pour Trino distinct du cluster Spark et d'inclure Iceberg v0.11 sur ce cluster.
- Lorsque vous utilisez AWS Glue comme catalogue pour Iceberg, assurez-vous que la base de données dans laquelle vous créez une table existe dans AWS Glue. Si vous utilisez des services tels que AWS Lake Formation et que vous ne parvenez pas à charger le catalogue, assurez-vous de disposer d'un accès approprié au service pour exécuter la commande.

Considérations relatives à l'utilisation d'Iceberg avec Flink

Lorsque vous utilisez AWS Glue comme catalogue pour Iceberg, assurez-vous que la base de données dans laquelle vous créez une table existe dans AWS Glue. Si vous utilisez des services tels que AWS Lake Formation et que vous ne parvenez pas à charger le catalogue, assurez-vous de disposer d'un accès approprié au service pour exécuter la commande.

Considérations relatives à l'utilisation d'Iceberg avec Hive

- Iceberg prend en charge les types de requêtes suivants :
 - Create table
 - Drop table
 - Insérer dans la table
 - Lire la table

- Seul le moteur d'exécution MR (MapReduce) est pris en charge pour les opérations DML (langage de manipulation de données), et le MR est obsolète dans Hive 3.1.3.
- AWS Glue Data Catalog n'est actuellement pas compatible avec Iceberg with Hive.
- La gestion des erreurs n'est pas suffisamment robuste. En cas de mauvaise configuration, les insertions dans les requêtes peuvent se terminer correctement. Cependant, le fait de ne pas mettre à jour les métadonnées peut entraîner des pertes de données.

Historique des versions d'Iceberg

Le tableau suivant répertorie la version d'Iceberg incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section [Version des composants de votre version](#) dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur les versions d'Iceberg

Étiquette Amazon EMR Release	Version Iceberg	Composants installés avec Iceberg
emr-7.2.0	1,5.0-amzn-0	Not available.
emr-7.1.0	1.4.3-amzn-0	Not available.
emr-7.0.0	1.4.2-amzn-0	Not available.
emr-6.15.0	1.4.0-amzn-0	Not available.
emr-6.14.0	1.3.1-amzn-0	Not available.
emr-6.13.0	1.3.0-amzn-1	Not available.
emr-6.12.0	1.3.0-amzn-0	Not available.
emr-6.11.1	1.2.0-amzn-0	Not available.
emr-6.11.0	1.2.0-amzn-0	Not available.

Étiquette Amazon EMR Release	Version Iceberg	Composants installés avec Iceberg
emr-6.10.1	1.1.0-amzn-0	Not available.
emr-6.10.0	1.1.0-amzn-0	Not available.
emr-6.9.1	0.14.1-amzn-0	Not available.
emr-6.9.0	0.14.1-amzn-0	Not available.
emr-6.8.1	0.14.0-amzn-0	Not available.
emr-6.8.0	0.14.0-amzn-0	Not available.
emr-6.7.0	0.13.1-amzn-0	Not available.
emr-6.6.0	0.13.1	Not available.
emr-6.5.0	0.12.0	Not available.

Notes de mise à jour d'Iceberg par version

- [Amazon EMR 6.9.0 - Notes de mise à jour d'Iceberg](#)

Amazon EMR 6.9.0 - Notes de mise à jour d'Iceberg

Amazon EMR 6.9.0 - Changements apportés par Iceberg

Type	Description
Fonctionnalité	Intégration d'Amazon EMR Flink à Iceberg.
Fonctionnalité	Intégration d'Amazon EMR Hive à Iceberg.
Fonctionnalité	Support pour mettre en cache les fichiers de métadonnées Iceberg sur Amazon FSx pour Lustre afin d'améliorer le temps de planification des requêtes.

Type	Description
Rétroportage	PR 5050 : Flink 1.15 : Support des options d'écriture dans les commentaires d'insertion en ligne. SQL
Rétroportage	PR 5282 AWS : Corrigez les échecs de PUT nouvelle tentative en ouvrant de nouveaux flux de fichiers de données.
Rétroportage	PR 5318 : Flink 1.15 : Comblez le fossé entre FlinkSource et IcebergSource (FLIP-27) et ajoutez une configuration opt-in pour utiliser FLIP la source -27 dans Flink. SQL
Rétroportage	PR 5344 : Flink 1.14 : Comblez le fossé entre FlinkSource et IcebergSource (FLIP-27) et ajoutez une configuration opt-in pour utiliser FLIP la source -27 dans Flink. SQL
Rétroportage	PR 5393 : Flink 1.14, 1.15 : évitez de convertir les métriques Iceberg en Flink dans le lecteur MetricContext de source -27. FLIP
Rétroportage	PR 5401 : Flink 1.14, 1.15 : IcebergSourceReader groupe manqué dans PR #5393 pour FLIP -27 métriques de lecteurs de sources.
Rétroportage	PR 5679 : Spark 3.2, 3.3 : correction de la propagation de nullabilité pour le nœud. MergeRows
Rétroportage	PR 5860 : Spark 3.3 : Correctif QueryFailure lors de l'exécution RewriteManifestProcedure sur des tables partitionnées par Date.
Rétroportage	PR 5880 : Spark 3.3 : correction de la nullité dans les projections. merge-on-read

Type	Description
Rétroportage	PR 5917 : Spark 3.2 : correction de la nullité dans les projections. merge-on-read

Bloc-notes Jupyter sur Amazon EMR

Le [bloc-notes Jupyter](#) est une application Web open source que vous pouvez utiliser pour créer et partager des documents qui contiennent du code interactif, des équations, des visualisations, et des textes narratifs. Amazon EMR vous offre trois options pour travailler avec les blocs-notes Jupyter :

Rubriques

- [EMR Studio](#)
- [Blocs-notes EMR basé sur Jupyter Notebook](#)
- [JupyterHub](#)

EMR Studio

Amazon EMR Studio est un environnement de développement intégré (IDE) basé sur le Web pour les [blocs-notes Jupyter](#) entièrement gérés qui s'exécutent sur des clusters Amazon EMR. Vous pouvez configurer un studio EMR pour que votre équipe développe, visualise et débogue des applications écrites en R, Python, Scala et PySpark.

Nous vous recommandons d'utiliser EMR Studio lorsque vous utilisez des blocs-notes Jupyter sur Amazon EMR. Pour plus d'informations, consultez [EMR Studio](#) dans le Guide de gestion d'Amazon EMR.

Blocs-notes EMR basé sur Jupyter Notebook

Blocs-notes EMR est un environnement [Bloc-notes Jupyter](#) intégré à la console Amazon EMR qui vous permet de créer rapidement des blocs-notes Jupyter, de les associer à des clusters Spark, puis d'ouvrir l'éditeur Jupyter Notebook dans la console pour exécuter des requêtes et du code à distance. Un bloc-notes EMR est enregistré dans Amazon S3 indépendamment des clusters pour un stockage durable, un accès rapide et une flexibilité accrue. Vous pouvez ouvrir plusieurs blocs-notes, associer plusieurs blocs-notes à un seul cluster et réutiliser un bloc-notes sur différents clusters.

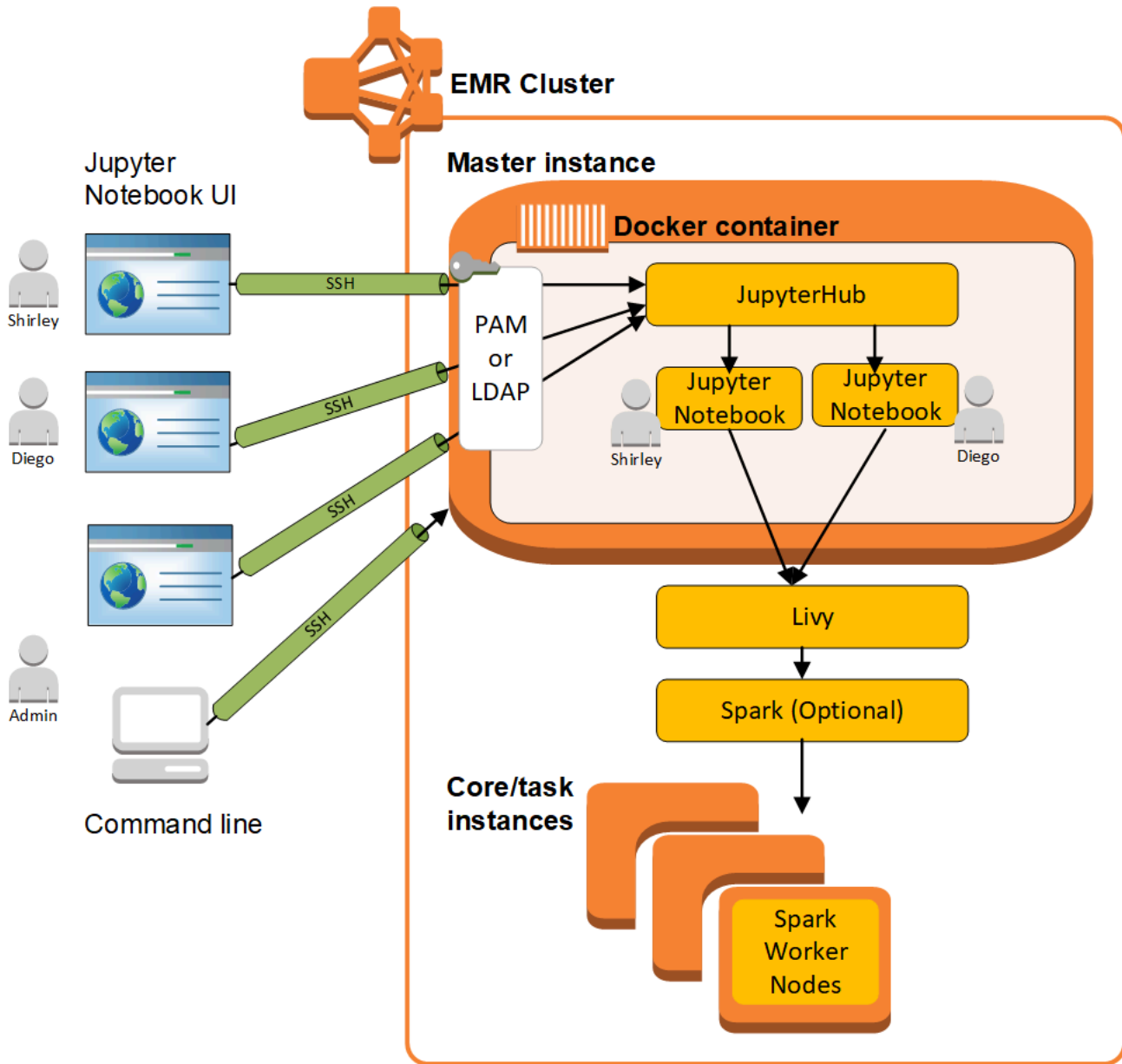
Pour plus d'informations, consultez [Blocs-notes EMR](#) dans le Guide de gestion d'Amazon EMR.

JupyterHub

Le [bloc-notes Jupyter](#) est une application Web open source que vous pouvez utiliser pour créer et partager des documents qui contiennent du code interactif, des équations, des visualisations, et des textes narratifs. [JupyterHub](#) vous permet d'héberger plusieurs instances d'un serveur de bloc-notes Jupyter mono-utilisateur. Lorsque vous créez un cluster avec JupyterHub, Amazon EMR crée un conteneur Docker sur le nœud principal du cluster. JupyterHub, tous les composants requis pour Jupyter et [Sparkmagic](#) fonctionnent dans le conteneur.

Sparkmagic est une bibliothèque de noyaux qui permet aux blocs-notes Jupyter d'interagir avec Apache [Spark exécuté](#) sur Amazon EMR via [Apache Livy](#) un serveur pour Spark. REST Spark et Apache Livy sont installés automatiquement lorsque vous créez un cluster avec JupyterHub. Le noyau Python 3 par défaut pour Jupyter est disponible, ainsi que les noyaux PySpark 3 PySpark, et Spark disponibles avec Sparkmagic. Vous pouvez utiliser ces noyaux pour exécuter du code Spark ad hoc et des SQL requêtes interactives à l'aide de Python et Scala. Vous pouvez installer manuellement des noyaux supplémentaires dans le conteneur Docker. Pour plus d'informations, consultez [Installation de noyaux et de bibliothèques supplémentaires](#).

Le schéma suivant décrit les composants d'Amazon EMR avec JupyterHub les méthodes d'authentification correspondantes pour les utilisateurs d'ordinateurs portables et l'administrateur. Pour plus d'informations, consultez [Ajout d'utilisateurs et d'administrateurs de bloc-notes Jupyter](#).



Le tableau suivant répertorie la version JupyterHub incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants avec lesquels Amazon EMR installe JupyterHub

Pour la version des composants installés JupyterHub dans cette version, voir [Versions des composants de la version 7.2.0.](#)

JupyterHub informations de version pour emr-7.2.0

Étiquette EMR de sortie Amazon	JupyterHub Version	Composants installés avec JupyterHub
emr-7.2.0	JupyterHub 1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Le tableau suivant répertorie la version JupyterHub incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants avec lesquels Amazon EMR installe JupyterHub

Pour connaître la version des composants installés JupyterHub dans cette version, consultez la [version 6.15.0 Versions des composants](#).

JupyterHub informations de version pour emr-6.15.0

Étiquette EMR de sortie Amazon	JupyterHub Version	Composants installés avec JupyterHub
emr-6.15.0	JupyterHub 1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Étiquette EMR de sortie Amazon	JupyterHub Version	Composants installés avec JupyterHub
		, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Le tableau suivant répertorie la version JupyterHub incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants avec lesquels Amazon EMR installe JupyterHub

Pour la version des composants installés JupyterHub dans cette version, voir la [version 5.36.2 Versions des composants](#).

JupyterHub informations de version pour emr-5.36.2

Étiquette EMR de sortie Amazon	JupyterHub Version	Composants installés avec JupyterHub
emr-5.36.2	JupyterHub 1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Le noyau Python 3 inclus dans Amazon EMR est JupyterHub la version 3.6.4.

Les bibliothèques installées dans le jupyterhub conteneur peuvent varier entre les EMR versions d'Amazon et les EC2 AMI versions d'Amazon.

Pour lister les bibliothèques installées à l'aide de **conda**

- Exécutez la commande suivante sur la ligne de commande du nœud principal :

```
sudo docker exec jupyterhub bash -c "conda list"
```

Pour lister les bibliothèques installées à l'aide de **pip**

- Exécutez la commande suivante sur la ligne de commande du nœud principal :

```
sudo docker exec jupyterhub bash -c "pip freeze"
```

Rubriques

- [Créez un cluster avec JupyterHub](#)
- [Considérations à prendre en compte lors de l'utilisation JupyterHub sur Amazon EMR](#)
- [Configuration JupyterHub](#)
- [Configuration de la persistance pour les blocs-notes dans Amazon S3](#)
- [Connexion au nœud principal et aux serveurs de bloc-notes](#)
- [JupyterHub configuration et administration](#)
- [Ajout d'utilisateurs et d'administrateurs de bloc-notes Jupyter](#)
- [Installation de noyaux et de bibliothèques supplémentaires](#)
- [JupyterHub historique des versions](#)

Créez un cluster avec JupyterHub

Vous pouvez créer un EMR cluster Amazon en JupyterHub utilisant le AWS Management Console AWS Command Line Interface, ou l'Amazon EMR API. Assurez-vous que le cluster n'est pas créé avec l'option d'arrêt automatique après exécution des étapes (option `--auto-terminate` dans l'AWS CLI). Assurez-vous également que les administrateurs et les utilisateurs du bloc-notes peuvent accéder à la paire de clés utilisée lors de la création du cluster. Pour plus d'informations, consultez la section [Utiliser une paire de clés pour les SSH informations d'identification](#) dans le guide EMR de gestion Amazon.

Création d'un cluster à JupyterHub l'aide de la console

Utilisez la procédure suivante pour créer un cluster JupyterHub installé à l'aide des options avancées de la EMR console Amazon.

Pour créer un EMR cluster Amazon JupyterHub installé à l'aide de la EMR console Amazon

1. Accédez à la nouvelle EMR console Amazon et sélectionnez **Basculer vers l'ancienne console** dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).
2. Choisissez **Créer un cluster** et **Go to advanced options** (Aller aux options avancées).
3. Sous **Software Configuration** (Configuration logicielle) :
 - Pour **Release**, sélectionnez **emr-5.36.2**, puis choisissez **JupyterHub**
 - Si vous utilisez **Spark**, pour utiliser le catalogue de données **AWS Glue** comme métastore pour **SparkSQL**, sélectionnez **Utiliser pour les métadonnées de table Spark**. Pour plus d'informations, consultez [Utiliser le catalogue de données AWS Glue comme métastore pour Spark SQL](#).
 - Pour **Modifier les paramètres du logiciel**, choisissez **Enter la configuration** et spécifiez des valeurs, ou choisissez **Load JSON from S3** et spécifiez un fichier **JSON** de configuration. Pour plus d'informations, consultez [Configuration JupyterHub](#).
4. Sous **Add steps (optional)** (Ajouter des étapes (facultatif)), configurez les étapes à exécuter lorsque le cluster est créé, assurez-vous que **Auto-terminate cluster after the last step is completed** (Arrêter automatiquement le cluster après l'exécution de la dernière étape) n'est pas sélectionné, puis choisissez **Next** (Suivant).
5. Choisissez les options **Hardware Configuration** (Configuration matérielle) et **Next** (Suivant). Pour plus d'informations, consultez [Configurer le matériel et le réseau du cluster](#) dans le Amazon EMR Management Guide.
6. Choisissez les options pour **General Cluster Settings** (Paramètres de cluster généraux) et **Next** (Suivant).
7. Choisissez **Security Options** (Options de sécurité) en spécifiant une paire de clés, puis choisissez **Create Cluster** (Créer un cluster).

Créez un cluster à JupyterHub l'aide du AWS CLI

Pour lancer un cluster avec JupyterHub, utilisez la `aws emr create-cluster` commande et, pour l'`--applicationsoption`, spécifiez `Name=JupyterHub`. L'exemple suivant lance un JupyterHub cluster sur Amazon EMR avec deux EC2 instances (une instance principale et une instance principale). En outre, le débogage est activé et les journaux sont stockés dans l'emplacement Amazon S3, comme spécifié par `--log-uri`. La paire de clés spécifiée permet d'accéder aux EC2 instances Amazon du cluster.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \  
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \  
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes  
KeyName=MyKeyPair
```

Considérations à prendre en compte lors de l'utilisation JupyterHub sur Amazon EMR

Tenez compte des points suivants lors de l'utilisation JupyterHub sur AmazonEMR.

Warning

Les blocs-notes utilisateur et les fichiers sont enregistrées dans le système de fichiers sur le nœud principal. Il s'agit d'un stockage éphémère qui n'est pas conservé en cas d'arrêt du cluster. Lors de l'arrêt du cluster, ces données sont perdues si elles ne sont pas sauvegardées. Nous vous recommandons de planifier des sauvegardes régulières en utilisant des tâches `cron` ou tout autre méthode adaptée à votre application.

De plus, les changements de configuration effectués dans le conteneur peuvent ne pas être conservés lors du redémarrage de celui-ci. Nous vous recommandons de créer des scripts ou d'automatiser de quelque manière que ce soit les configurations de conteneur, de sorte que vous puissiez reproduire les personnalisations plus rapidement.

- L'authentification Kerberos configurée à l'aide d'une configuration de EMR sécurité Amazon n'est pas prise en charge.
- [OAuthenticator](#) n'est pas pris en charge.

Configuration JupyterHub

Vous pouvez personnaliser la configuration des JupyterHub blocs-notes d'Amazon EMR et d'utilisateurs individuels en vous connectant au nœud principal du cluster et en modifiant les fichiers de configuration. Après avoir modifié les valeurs, redémarrez le conteneur `jupyterhub`.

Modifiez les propriétés dans les fichiers suivants pour configurer JupyterHub des blocs-notes Jupyter individuels :

- `jupyterhub_config.py` – Par défaut, ce fichier est enregistré dans le répertoire `/etc/jupyter/conf/` sur le nœud principal. Pour plus d'informations, consultez la section [Principes de base de la configuration](#) dans la JupyterHub documentation.
- `jupyter_notebook_config.py` – Ce fichier est enregistré dans le répertoire `/etc/jupyter/` par défaut et copié dans le conteneur `jupyterhub` en tant que valeur par défaut. Pour plus d'informations, consultez [Fichier de configuration et options de ligne de commande](#) dans la documentation sur les bloc-notes Jupyter.

Vous pouvez également utiliser la classification de configuration `jupyter-sparkmagic-conf` pour personnaliser Sparkmagic, ce qui met à jour les valeurs du fichier `config.json` pour Sparkmagic. Pour plus d'informations sur les paramètres disponibles, consultez le fichier [example_config.json](#) sur GitHub. Pour plus d'informations sur l'utilisation des classifications de configuration avec les applications d'AmazonEMR, consultez [Configuration des applications](#).

L'exemple suivant lance un cluster en utilisant le AWS CLI, en référençant le fichier `MyJupyterConfig.json` pour les paramètres de classification de configuration de Sparkmagic.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --use-default-roles --release-label emr-5.14.0 \  
--applications Name=Jupyter --instance-type m4.xlarge --instance-count 3 \  
--ec2-attributes KeyName=MyKey,SubnetId=subnet-1234a5b6 --configurations file://\  
MyJupyterConfig.json
```

Exemples de contenu du fichier MyJupyterConfig.json :

```
[  
  {  
    "Classification": "jupyter-sparkmagic-conf",  
    "Properties": {  
      "kernel_python_credentials" : "{\"username\":\"diego\", \"base64_password\":  
\"mypass\", \"url\":\"http://localhost:8998\", \"auth\":\"None\"}"  
    }  
  }  
]
```

Note

Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier des classifications de configuration supplémentaires pour chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, utilisez la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).

Configuration de la persistance pour les blocs-notes dans Amazon S3

Vous pouvez configurer un JupyterHub cluster dans Amazon EMR afin que les blocs-notes enregistrés par un utilisateur soient conservés dans Amazon S3, en dehors du stockage éphémère sur les instances de cluster. EC2

Vous spécifiez une persistance Amazon S3 à l'aide de la classification de configuration `jupyter-s3-conf` lorsque vous créez un cluster. Pour plus d'informations, consultez [Configuration des applications](#).

En plus de l'activation de persistance Amazon S3 à l'aide de la propriété `s3.persistence.enabled`, vous devez spécifier un compartiment dans Amazon S3 où les blocs-

notes sont enregistrés à l'aide de la propriété `s3.persistence.bucket`. Les blocs-notes de chaque utilisateur sont enregistrés dans un dossier `jupyter/jupyterhub-user-name` dans le compartiment spécifié. Le compartiment doit déjà exister dans Amazon S3, et le rôle du profil d'EC2instance que vous spécifiez lors de la création du cluster doit disposer des autorisations d'accès au compartiment (par défaut, le rôle est `EMR_EC2_DefaultRole`). Pour plus d'informations, consultez [Configurer IAM les rôles pour les EMR autorisations Amazon relatives aux AWS services](#).

Lorsque vous lancez un nouveau cluster à l'aide des mêmes propriétés de classification de configuration, les utilisateurs peuvent ouvrir des blocs-notes avec le contenu à partir de l'emplacement enregistré.

Notez que lorsque vous importez des fichiers sous forme de modules dans un bloc-notes lorsque Amazon S3 est activé, cela entraîne le chargement des fichiers sur Amazon S3. Lorsque vous importez des fichiers sans activer la persistance Amazon S3, ils sont chargés JupyterHub dans votre conteneur.

L'exemple suivant active la persistance Amazon S3. Les cahiers enregistrés par les utilisateurs sont enregistrés dans le dossier `s3://MyJupyterBackups/jupyter/jupyterhub-user-name` pour chaque utilisateur, où `jupyterhub-user-name` est un nom d'utilisateur, par exemple `diego`.

```
[
  {
    "Classification": "jupyter-s3-conf",
    "Properties": {
      "s3.persistence.enabled": "true",
      "s3.persistence.bucket": "MyJupyterBackups"
    }
  }
]
```

Connexion au nœud principal et aux serveurs de bloc-notes

JupyterHub les administrateurs et les utilisateurs d'ordinateurs portables doivent se connecter au nœud principal du cluster via un SSH tunnel, puis se connecter aux interfaces Web desservies par JupyterHub le nœud principal. Pour plus d'informations sur la configuration d'un SSH tunnel et son utilisation pour les connexions Web par proxy, consultez [Connect to the cluster](#) dans le Amazon EMR Management Guide.

Par défaut, JupyterHub Amazon EMR est disponible via le port 9443 sur le nœud principal. Le JupyterHub proxy interne dessert également les instances de bloc-notes via le port 9443. JupyterHub et les interfaces Web de Jupyter sont accessibles à l'URL aide d'un modèle suivant :

```
https://MasterNodeDNS:9443
```

Vous pouvez spécifier un autre port à l'aide de la propriété `c.JupyterHub.port` dans le fichier `jupyterhub_config.py`. Pour plus d'informations, consultez la section [Notions de base de la mise en réseau](#) dans la JupyterHub documentation.

Par défaut, JupyterHub Amazon EMR utilise un certificat auto-signé pour le SSL chiffré à l'aide de HTTPS. Les utilisateurs sont invités à approuver le certificat auto-signé lorsqu'ils se connectent. Vous pouvez utiliser un certificat approuvé et des clés qui vous sont propres. Remplacez le fichier de certificat par défaut, `server.crt` et le fichier de clé `server.key` dans le répertoire `/etc/jupyter/conf/` du nœud principal par un certificat et des fichiers de clé qui vous sont propres. Utilisez les propriétés `c.JupyterHub.ssl_cert` et `c.JupyterHub.ssl_key` du fichier `jupyterhub_config.py` pour spécifier vos SSL matériaux. Pour plus d'informations, consultez [la section Paramètres de sécurité](#) dans la JupyterHub documentation. Après avoir mis à jour `jupyterhub_config.py`, redémarrez le conteneur.

JupyterHub configuration et administration

JupyterHub et les composants associés s'exécutent dans un conteneur Docker nommé `jupyterhub` qui exécute le système d'exploitation Ubuntu. Il existe plusieurs méthodes pour administrer des composants s'exécutant à l'intérieur du conteneur.

Warning

Les personnalisations que vous effectuez dans le conteneur peuvent ne pas être conservées si le conteneur redémarre. Nous vous recommandons de créer des scripts ou d'automatiser de quelque manière que ce soit les configurations de conteneur, de sorte que vous puissiez reproduire les personnalisations plus rapidement.

Administration à l'aide de la ligne de commande

Lorsque vous êtes connecté au nœud principal à l'aide de SSH, vous pouvez émettre des commandes à l'aide de l'interface de ligne de commande Docker (CLI) et en spécifiant le conteneur

par son nom (`jupyterhub`) ou son ID. Par exemple, `sudo docker exec jupyterhub command` exécute des commandes reconnues par le système d'exploitation ou par une application exécutée dans le conteneur. Vous pouvez utiliser cette méthode pour ajouter des utilisateurs au système d'exploitation et installer des applications et des bibliothèques supplémentaires au sein du conteneur Docker. Par exemple, l'image de conteneur par défaut inclut Conda pour l'installation de package ; vous pouvez donc exécuter la commande suivante sur la ligne de commande du nœud principal pour installer une application, Keras, dans le conteneur :

```
sudo docker exec jupyterhub conda install keras
```

Administration par soumission d'étapes

Les étapes constituent un moyen de soumettre le travail à un cluster. Vous pouvez soumettre des étapes lors du lancement d'un cluster ou lorsque celui-ci est en cours d'exécution. Les commandes exécutées sur la ligne de commande peuvent être soumises sous forme d'étapes à l'aide de `command-runner.jar`. Pour plus d'informations, consultez [Work with steps using the CLI and console](#) dans le Amazon EMR Management Guide et [Exécuter des commandes et des scripts sur un cluster Amazon EMR](#).

Par exemple, vous pouvez utiliser la AWS CLI commande suivante sur un ordinateur local pour installer Keras de la même manière que vous l'avez fait à partir de la ligne de commande du nœud principal dans l'exemple précédent :

```
aws emr add-steps --cluster-id MyClusterID --steps Name="Command Runner",Jar="command-runner.jar",Args="/usr/bin/sudo","/usr/bin/docker","exec","jupyterhub","conda","install","keras"
```

De plus, vous pouvez créer un script avec une séquence d'étapes, le charger sur Amazon S3, puis utiliser `script-runner.jar` pour l'exécuter lors de la création du cluster ou l'ajouter en tant qu'étape. Pour plus d'informations, consultez [Exécuter des commandes et des scripts sur un cluster Amazon EMR](#). Pour obtenir un exemple, consultez [the section called "Exemple : script bash pour ajouter plusieurs utilisateurs"](#).

Administration à l'aide de REST APIs

Jupyter JupyterHub, et le HTTP proxy pour JupyterHub fournir REST APIs que vous pouvez utiliser pour envoyer des demandes. Pour envoyer des demandes à JupyterHub, vous devez transmettre un API jeton avec la demande. Vous pouvez utiliser la `curl` commande depuis la ligne de commande

du nœud principal pour exécuter REST des commandes. Pour plus d'informations, consultez les ressources suivantes :

- [Les JupyterHub utilisations figurent REST API](#) dans la documentation de JupyterHub, qui inclut des instructions pour générer API des jetons
- [Serveur Jupyter Notebook activé API](#) GitHub
- [configurable-http-proxy](#) sur GitHub

L'exemple suivant montre comment utiliser le REST API for JupyterHub pour obtenir une liste d'utilisateurs. La commande transmet un jeton d'administration généré précédemment et utilise le port par défaut, 9443 JupyterHub, pour rediriger la sortie vers [jq afin de faciliter](#) la visualisation :

```
curl -XGET -s -k https://$HOST:9443/hub/api/users \  
-H "Authorization: token $admin_token" | jq .
```

Ajout d'utilisateurs et d'administrateurs de bloc-notes Jupyter

Vous pouvez utiliser l'une des deux méthodes permettant aux utilisateurs de s'authentifier JupyterHub afin qu'ils puissent créer des blocs-notes et, éventuellement, les administrer. JupyterHub La méthode la plus simple consiste à utiliser JupyterHub le module d'authentification enfichable ()PAM. En outre, JupyterHub Amazon EMR prend en charge le [plug-in d'LDAPauthentification JupyterHub permettant](#) d'obtenir l'identité des utilisateurs à partir d'un LDAP serveur, tel qu'un serveur Microsoft Active Directory. Des instructions et des exemples pour l'ajout d'utilisateurs avec chaque méthode d'authentification sont fournis dans cette section.

JupyterHub sur Amazon EMR possède un utilisateur par défaut doté d'autorisations d'administrateur. Le nom d'utilisateur est `jovyan`, et le mot de passe est `jupyter`. Nous vous recommandons vivement de remplacer cet utilisateur par un autre utilisateur disposant des autorisations d'administrateur. Vous pouvez effectuer cette opération via une étape lorsque vous créez le cluster, ou en vous connectant au nœud maître lorsque le cluster est en cours d'exécution.

Rubriques

- [Utilisation de l'PAMauthentification](#)
- [Utilisation de l'LDAPauthentification](#)
- [Emprunt d'identité de l'utilisateur](#)

Utilisation de l'PAMauthentification

La création d'PAMutilisateurs JupyterHub sur Amazon EMR est un processus en deux étapes. La première étape consiste à ajouter des utilisateurs au système d'exploitation qui s'exécute dans le conteneur `jupyterhub` sur le nœud principal, et d'ajouter le répertoire de base correspondant pour chaque utilisateur. La deuxième étape consiste à ajouter ces utilisateurs du système d'exploitation en tant qu' `JupyterHubutilisateurs`, un processus connu sous le nom de liste blanche dans `JupyterHub`. Une fois qu'un `JupyterHub` utilisateur est ajouté, il peut se connecter au `JupyterHub` URL et fournir ses informations d'identification de système d'exploitation pour y accéder.

Lorsqu'un utilisateur se connecte, `JupyterHub` ouvre l'instance de serveur de bloc-notes pour cet utilisateur, qui est enregistrée dans le répertoire personnel de l'utilisateur sur le nœud principal, à savoir `/var/lib/jupyter/home/username`. Si aucune instance de serveur de bloc-notes n'existe, `JupyterHub` génère une instance de bloc-notes dans le répertoire personnel de l'utilisateur. Les sections suivantes montrent comment ajouter des utilisateurs individuellement au système d'exploitation et à `JupyterHub`, puis comment utiliser un script bash rudimentaire qui ajoute plusieurs utilisateurs.

Ajout d'un utilisateur du système d'exploitation au conteneur

L'exemple suivant utilise tout d'abord la commande [useradd](#) dans le conteneur pour ajouter un utilisateur unique, `diego`, et créer un répertoire de base pour cet utilisateur. La deuxième commande utilise [chpasswd](#) pour établir le mot de passe associé à `diego` pour cet utilisateur. Les commandes sont exécutées sur la ligne de commande du nœud maître lorsque vous êtes connecté à l'aide de SSH. Vous pouvez également exécuter ces commandes à l'aide d'une étape, comme décrit précédemment dans [Administration par soumission d'étapes](#).

```
sudo docker exec jupyterhub useradd -m -s /bin/bash -N diego
sudo docker exec jupyterhub bash -c "echo diego:diego | chpasswd"
```

Ajouter un JupyterHub utilisateur

Vous pouvez utiliser le panneau d'administration dans `JupyterHub` ou REST API pour ajouter des utilisateurs et des administrateurs, ou simplement des utilisateurs.

Pour ajouter des utilisateurs et des administrateurs à l'aide du panneau d'administration dans `JupyterHub`

1. Connectez-vous au nœud principal en utilisant SSH et connectez-vous à `https://MasterNodeDNS:9443` avec une identité dotée d'autorisations d'administrateur.

2. Choisissez Control Panel (Panneau de configuration), Admin.
3. Choisissez User (Utilisateur), Add Users (Ajouter des utilisateurs), ou choisissez Admin, Add Admins (Ajouter des administrateurs).

Pour ajouter un utilisateur à l'aide du REST API

1. Connectez-vous au nœud principal en utilisant SSH et utilisez la commande suivante sur le nœud principal, ou exécutez la commande en tant qu'étape par étape.
2. Acquérir un jeton administratif pour effectuer API des demandes et le remplacer *AdminToken* à l'étape suivante avec ce jeton.
3. Utilisez la commande suivante, en remplaçant *UserName* avec un utilisateur du système d'exploitation créé dans le conteneur.

```
curl -XPOST -H "Authorization: token AdminToken" "https://$(hostname):9443/hub/api/users/UserName"
```

Note

Vous êtes automatiquement ajouté en tant qu'utilisateur JupyterHub non administrateur lorsque vous vous connectez à l'interface JupyterHub Web pour la première fois.

Exemple : script bash pour ajouter plusieurs utilisateurs

L'exemple de script bash suivant associe les étapes précédentes de cette section pour créer plusieurs JupyterHub utilisateurs. Le script peut être exécuté directement sur le nœud principal, ou chargé sur Amazon S3 et exécuté en tant qu'étape.

Le script établit d'abord un tableau de noms d'utilisateur et utilise la `jupyterhub token` commande pour créer un API jeton pour l'administrateur par défaut, `jovyan`. Ensuite, il crée un utilisateur du système d'exploitation dans le conteneur `jupyterhub` pour chaque utilisateur, et lui affecte un mot de passe initial équivalent à son nom d'utilisateur. Enfin, il appelle l'opération de création de chaque utilisateur dans JupyterHub. Il transmet le jeton généré plus tôt dans le script et redirige la REST réponse vers celui-ci `jq` pour en faciliter la visualisation.

```
# Bulk add users to container and JupyterHub with temp password of username
```

```

set -x
USERS=(shirley diego ana richard li john mary anaya)
TOKEN=$(sudo docker exec jupyterhub /opt/conda/bin/jupyterhub token jovyan | tail -1)
for i in "${USERS[@]}";
do
    sudo docker exec jupyterhub useradd -m -s /bin/bash -N $i
    sudo docker exec jupyterhub bash -c "echo $i:$i | chpasswd"
    curl -XPOST --silent -k https://$(hostname):9443/hub/api/users/$i \
-H "Authorization: token $TOKEN" | jq
done

```

Enregistrez le script dans un emplacement Amazon S3 tel que `s3://mybucket/createjupyterusers.sh`. Vous pouvez ensuite utiliser `script-runner.jar` pour l'exécuter en tant qu'étape.

Exemple : exécution du script lors de la création d'un cluster (AWS CLI)

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```

aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes
KeyName=MyKeyPair \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
mybucket/createjupyterusers.sh"]

```

Exécution du script sur un cluster existant (AWS CLI)

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr add-steps --cluster-id j-XXXXXXXX --steps Type=CUSTOM_JAR,\
Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://
mybucket/createjupyterusers.sh"]
```

Utilisation de l'LDAP authentication

Le Lightweight Directory Access Protocol (LDAP) est un protocole d'application permettant d'interroger et de modifier des objets correspondant à des ressources telles que des utilisateurs et des ordinateurs stockés dans un fournisseur de services d'annuaire LDAP compatible tel qu'Active Directory ou un serveur Open LDAP Server. Vous pouvez utiliser le [plugin LDAP d'authentification pour JupyterHub](#) with JupyterHub sur Amazon EMR pour l'authentification LDAP des utilisateurs. Le plugin gère les sessions de connexion des LDAP utilisateurs et fournit des informations utilisateur à Jupyter. Cela permet aux utilisateurs de se connecter à JupyterHub des blocs-notes en utilisant les informations d'identification de leur identité stockées sur un LDAP serveur compatible.

Les étapes décrites dans cette section vous guident à travers les étapes suivantes pour configurer et activer LDAP l'utilisation du plug-in LDAP Authenticator pour JupyterHub. Pour effectuer ces étapes, vous devez être connecté à la ligne de commande du nœud principal. Pour plus d'informations, consultez [Connexion au nœud principal et aux serveurs de bloc-notes](#).

1. Créez un fichier LDAP de configuration contenant des informations sur le LDAP serveur, telles que l'adresse IP de l'hôte, le port, les noms de liaison, etc.
2. Modifiez `/etc/jupyter/conf/jupyterhub_config.py` pour activer le plug-in LDAP d'authentification pour JupyterHub
3. Créez et exécutez un script qui se configure LDAP dans le jupyterhub conteneur.
4. Interrogez LDAP les utilisateurs, puis créez des répertoires personnels dans le conteneur pour chaque utilisateur. JupyterHub nécessite des répertoires personnels pour héberger les blocs-notes.
5. Exécutez un script qui redémarre JupyterHub

Important

Avant de procéder à la configuration LDAP, testez votre infrastructure réseau pour vous assurer que le LDAP serveur et le nœud principal du cluster peuvent communiquer selon les

besoins. TLS utilise généralement le port 389 via une TCP connexion simple. Si votre LDAP connexion l'utilise SSL, le TCP port connu pour SSL est 636.

Création du fichier LDAP de configuration

L'exemple ci-dessous utilise les valeurs de configuration d'espace réservé suivantes. Remplacez ces valeurs par des paramètres qui correspondent à votre implémentation.

- Le LDAP serveur exécute la version 3 et est disponible sur le port 389. Il s'agit du SSL non-port standard pour LDAP.
- Le nom unique de base est `dc=example, dc=org`.

Utilisez un éditeur de texte pour créer le fichier [ldap.conf](#), avec un contenu similaire à ce qui suit. Utilisez des valeurs adaptées à votre LDAP implémentation. Remplacez *host* avec l'adresse IP ou le nom d'hôte résoluble de votre LDAP serveur.

```
base dc=example,dc=org
uri ldap://host
ldap_version 3
binddn cn=admin,dc=example,dc=org
bindpw admin
```

Activer le LDAP plug-in d'authentification pour JupyterHub

Utilisez un éditeur de texte pour modifier le fichier `/etc/jupyter/conf/jupyterhub_config.py` et ajouter les propriétés [ldapauthenticator](#) similaires à ce qui suit. Remplacez *host* avec l'adresse IP ou le nom d'hôte résoluble du LDAP serveur. L'exemple suppose que les objets utilisateur se trouvent au sein d'une unité organisationnelle (ou) nommée *people*, et utilise les composants de nom unique que vous avez définis précédemment à l'aide de `ldap.conf`.

```
c.JupyterHub.authenticator_class = 'ldapauthenticator.LDAPAuthenticator'
c.LDAPAuthenticator.use_ssl = False
c.LDAPAuthenticator.server_address = 'host'
c.LDAPAuthenticator.bind_dn_template = 'cn={username},ou=people,dc=example,dc=org'
```

Configurer LDAP dans le conteneur

Utilisez un éditeur de texte pour créer un script bash avec le contenu suivant :

```
#!/bin/bash

# Uncomment the following lines to install LDAP client libraries only if
# using Amazon EMR release version 5.14.0. Later versions install libraries by default.
# sudo docker exec jupyterhub bash -c "sudo apt-get update"
# sudo docker exec jupyterhub bash -c "sudo apt-get -y install libnss-ldap libpam-ldap
  ldap-utils nscd"

# Copy ldap.conf
sudo docker cp ldap.conf jupyterhub:/etc/ldap/
sudo docker exec jupyterhub bash -c "cat /etc/ldap/ldap.conf"

# configure nss switch
sudo docker exec jupyterhub bash -c "sed -i 's/^(^passwd.*\)/\1 ldap/g' /etc/
nsswitch.conf"
sudo docker exec jupyterhub bash -c "sed -i 's/^(^group.*\)/\1 ldap/g' /etc/
nsswitch.conf"
sudo docker exec jupyterhub bash -c "sed -i 's/^(^shadow.*\)/\1 ldap/g' /etc/
nsswitch.conf"
sudo docker exec jupyterhub bash -c "cat /etc/nsswitch.conf"

# configure PAM to create home directories
sudo docker exec jupyterhub bash -c "echo 'session required          pam_mkhome.so
  skel=/etc/skel umask=077' >> /etc/pam.d/common-session"
sudo docker exec jupyterhub bash -c "cat /etc/pam.d/common-session"

# restart nscd service
sudo docker exec jupyterhub bash -c "sudo service nscd restart"

# Test
sudo docker exec jupyterhub bash -c "getent passwd"

# Install ldap plugin
sudo docker exec jupyterhub bash -c "pip install jupyterhub-ldapauthenticator"
```

Enregistrez le script sur le nœud principal, puis exécutez-le à partir de la ligne de commande du nœud principal. Par exemple, avec le script enregistré sous le nom `configure_ldap_client.sh`, rendez le fichier exécutable :

```
chmod +x configure_ldap_client.sh
```

Exécutez ensuite le script :


```
./configure_ldap_client.sh
```

Ajout d'attributs à Active Directory

Pour rechercher chaque utilisateur et créer l'entrée appropriée dans la base de données, le conteneur JupyterHub docker nécessite les UNIX propriétés suivantes pour l'objet utilisateur correspondant dans Active Directory. Pour plus d'informations, consultez la section Comment continuer à modifier les attributsGID/UIDRFC2307 maintenant que le plug-in d'attributs Unix n'est plus disponible pour le composant logiciel enfichable Utilisateurs et ordinateurs MMC Active Directory ? dans l'article [Clarification concernant l'état de la gestion des identités pour Unix \(IDMU\) et le rôle NIS du serveur dans la version préliminaire technique de Windows Server 2016 et au-delà.](#)

- `homeDirectory`

Il s'agit de l'emplacement de l'annuaire de base de l'utilisateur, qui est généralement `/home/username`.

- `gidNumber`

Il s'agit d'une valeur supérieure à 60 000 qui n'est pas déjà utilisée par un autre utilisateur. Vérifiez le fichier `etc/passwd` pour les GID en cours d'utilisation.

- `uidNumber`

Il s'agit d'une valeur supérieure à 60 000 qui n'est pas déjà utilisée par un autre groupe. Vérifiez le fichier `etc/group` pour les UID en cours d'utilisation.

- `uid`

C'est la même chose que `username`.

Création de répertoires de base utilisateur

JupyterHub a besoin de répertoires personnels dans le conteneur pour authentifier LDAP les utilisateurs et stocker les données d'instance. L'exemple suivant montre deux utilisateurs, Shirley et Diego, dans le LDAP répertoire.

La première étape consiste à interroger le LDAP serveur pour obtenir les informations d'identifiant d'utilisateur et d'identifiant de groupe de chaque utilisateur à l'aide de [ldapsearch](#), comme indiqué dans l'exemple suivant, en remplaçant `host` avec l'adresse IP ou le nom d'hôte résoluble de votre LDAP serveur :

```
ldapsearch -x -H ldap://host \  
-D "cn=admin,dc=example,dc=org" \  
-w admin \  
-b "ou=people,dc=example,dc=org" \  
-s sub \  
"(objectclass=*)" uidNumber gidNumber
```

La `ldapsearch` commande renvoie une réponse LDIF au format -qui ressemble à ce qui suit pour les utilisateurs Shirley et Diego.

```
# extended LDIF  
  
# LDAPv3  
# base <ou=people,dc=example,dc=org> with scope subtree  
# filter: (objectclass=*)  
# requesting: uidNumber gidNumber sn  
  
# people, example.org  
dn: ou=people,dc=example,dc=org  
  
# diego, people, example.org  
dn: cn=diego,ou=people,dc=example,dc=org  
sn: B  
uidNumber: 1001  
gidNumber: 100  
  
# shirley, people, example.org  
dn: cn=shirley,ou=people,dc=example,dc=org  
sn: A  
uidNumber: 1002  
gidNumber: 100  
  
# search result  
search: 2  
result: 0 Success  
  
# numResponses: 4  
# numEntries: 3
```

À l'aide des informations contenues dans la réponse, exécutez des commandes dans le conteneur de manière à créer un répertoire de base pour chaque nom commun d'utilisateur (cn). Utilisez la valeur

de `uidNumber` et de `gidNumber` pour corriger la propriété du répertoire de base pour cet utilisateur. Les exemples de commandes suivants le font pour l'utilisateur : *shirley*.

```
sudo docker container exec jupyterhub bash -c "mkdir /home/shirley"
sudo docker container exec jupyterhub bash -c "chown -R $uidNumber /home/shirley"
sudo docker container exec jupyterhub bash -c "sudo chgrp -R $gidNumber /home/shirley"
```

Note

LDAPauthenticator for JupyterHub ne prend pas en charge la création d'utilisateurs locaux. Pour plus d'informations, consultez la [note de configuration de l'LDAPauthenticator sur la création d'utilisateurs locaux](#).

Pour créer manuellement un utilisateur local, utilisez la commande suivante.

```
sudo docker exec jupyterhub bash -c "echo 'shirley:x:$uidNumber:$gidNumber::/home/shirley:/bin/bash' >> /etc/passwd"
```

Redémarrer le JupyterHub conteneur

Exécutez la commande suivante pour redémarrer le conteneur `jupyterhub` :

```
sudo docker stop jupyterhub
sudo docker start jupyterhub
```

Emprunt d'identité de l'utilisateur

Une tâche Spark exécutée dans un bloc-notes Jupyter traverse plusieurs applications lors de son exécution sur Amazon. EMR Par exemple, PySpark 3 codes exécutés par un utilisateur dans Jupyter sont reçus par Sparkmagic, qui utilise une HTTP POST demande pour le soumettre à Livy, qui crée ensuite une tâche Spark à exécuter sur le cluster en utilisant. YARN

Par défaut, les YARN tâches soumises de cette manière sont exécutées en tant qu'utilisateur `livy`, quel que soit l'utilisateur à l'origine de la tâche. En configurant l'usurpation d'identité, vous pouvez faire en sorte que l'ID utilisateur de l'utilisateur du bloc-notes soit également l'utilisateur associé à la YARN tâche. Plutôt que les tâches soient lancées par les deux utilisateurs `shirley` et `diego` associés à l'utilisateur `livy`, les tâches initiées par chaque utilisateur sont associées respectivement à `shirley` et à `diego`. Cela vous permet de vérifier l'utilisation de Jupyter et de gérer les applications au sein de votre organisation.

Cette configuration est prise en charge uniquement lorsque des appels de Sparkmagic vers Livy ne sont pas authentifiés. Les applications qui fournissent une couche d'authentification ou de redirection entre les applications Hadoop et Livy (par exemple, Apache Knox Gateway) ne sont pas prises en charge. Les étapes de configuration de l'emprunt d'identité utilisateur décrites dans cette section supposent que Livy JupyterHub et Livy s'exécutent sur le même nœud principal. Si votre application possède des clusters distincts, elle [Étape 3 : créer des HDFS répertoires personnels pour les utilisateurs](#) doit être modifiée pour que les HDFS répertoires soient créés sur le nœud Livy Master.

Étapes pour configurer l'emprunt d'identité de l'utilisateur

- [Étape 1 : Configurer Livy](#)
- [Étape 2 : Ajouter des utilisateurs](#)
- [Étape 3 : créer des HDFS répertoires personnels pour les utilisateurs](#)

Étape 1 : Configurer Livy

Vous utilisez les classifications de configuration `livy-conf` et `core-site` lorsque vous créez un cluster pour activer l'emprunt d'identité de l'utilisateur Livy, comme illustré dans l'exemple suivant. Enregistrez la classification de configuration sous forme de JSON fichier, puis référencez-la lorsque vous créez le cluster, ou spécifiez la classification de configuration en ligne. Pour plus d'informations, consultez [Configuration des applications](#).

```
[
  {
    "Classification": "livy-conf",
    "Properties": {
      "livy.impersonation.enabled": "true"
    }
  },
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.livy.groups": "*",
      "hadoop.proxyuser.livy.hosts": "*"
    }
  }
]
```

Étape 2 : Ajouter des utilisateurs

Ajoutez des JupyterHub utilisateurs à l'aide de PAM ouLDAP. Pour plus d'informations, consultez [Utilisation de l'PAMauthentification](#) et [Utilisation de l'LDAPauthentification](#).

Étape 3 : créer des HDFS répertoires personnels pour les utilisateurs

Vous vous êtes connecté au nœud maître pour créer des utilisateurs. Toujours connecté au nœud maître, copiez le contenu ci-dessous et enregistrez-le dans un fichier script. Le script crée des HDFS répertoires personnels pour chaque JupyterHub utilisateur du nœud principal. Le script suppose que vous utilisez l'ID utilisateur administrateur par défaut, *jovyan*.

```
#!/bin/bash

CURL="curl --silent -k"
HOST=$(curl -s http://169.254.169.254/latest/meta-data/local-hostname)

admin_token() {
    local user=jovyan
    local pwd=jupyter
    local token=$(($CURL https://$HOST:9443/hub/api/authorizations/token \
        -d "{\"username\":\"$user\", \"password\":\"$pwd\"}" | jq ".token")
    if [[ $token != null ]]; then
        token=$(echo $token | sed 's//g')
    else
        echo "Unable to get Jupyter API Token."
        exit 1
    fi
    echo $token
}

# Get Jupyter Admin token
token=$(admin_token)

# Get list of Jupyter users
users=$(curl -XGET -s -k https://$HOST:9443/hub/api/users \
    -H "Authorization: token $token" | jq '.[].name' | sed 's//g')

# Create HDFS home dir
for user in ${users[@]};
do
    echo "Create hdfs home dir for $user"
    hadoop fs -mkdir /user/$user
```

```
hadoop fs -chmod 777 /user/$user
done
```

Installation de noyaux et de bibliothèques supplémentaires

Lorsque vous créez un cluster JupyterHub sur AmazonEMR, le noyau Python 3 par défaut pour Jupyter ainsi que les noyaux Spark PySpark et Spark pour Sparkmagic sont installés sur le conteneur Docker. Vous pouvez installer des noyaux supplémentaires. Vous pouvez également installer des bibliothèques et des packages supplémentaires, puis les importer pour le shell approprié.

Installation d'un noyau

Les noyaux sont installés dans le conteneur Docker. La manière la plus simple de faire cela consiste à créer un script bash avec des commandes d'installation, de l'enregistrer sur le nœud principal, puis d'utiliser la commande `sudo docker exec jupyterhub script_name` pour exécuter le script dans le conteneur `jupyterhub`. L'exemple de script suivant installe le noyau, puis installe quelques bibliothèques pour ce noyau sur le nœud principal afin que vous puissiez ultérieurement importer les bibliothèques à l'aide du noyau dans Jupyter.

```
#!/bin/bash

# Install Python 2 kernel
conda create -n py27 python=2.7 anaconda
source /opt/conda/envs/py27/bin/activate
apt-get update
apt-get install -y gcc
/opt/conda/envs/py27/bin/python -m pip install --upgrade ipykernel
/opt/conda/envs/py27/bin/python -m ipykernel install

# Install libraries for Python 2
/opt/conda/envs/py27/bin/pip install paramiko nltk scipy numpy scikit-learn pandas
```

Pour installer le noyau et les bibliothèques dans le conteneur, ouvrez une connexion terminale sur le nœud principal, enregistrez le script sous `/etc/jupyter/install_kernels.sh`, et exécutez la commande suivante sur la ligne de commande du nœud principal :

```
sudo docker exec jupyterhub bash /etc/jupyter/install_kernels.sh
```

Utilisation des bibliothèques et installation de bibliothèques supplémentaires

Un ensemble essentiel de bibliothèques d'apprentissage automatique et de science des données pour Python 3 est préinstallé JupyterHub sur AmazonEMR. Vous pouvez utiliser `sudo docker exec jupyterhub bash -c "conda list"` et `sudo docker exec jupyterhub bash -c "pip freeze"`.

Si une tâche Spark a besoin de bibliothèques sur les nœuds de travail, nous vous recommandons d'utiliser une action d'amorçage pour exécuter un script afin d'installer les bibliothèques lorsque vous créez le cluster. Les actions d'amorçage s'exécutent sur tous les nœuds du cluster pendant le processus de création du cluster, ce qui simplifie l'installation. Si vous installez des bibliothèques sur des nœuds principaux/de travail une fois que le cluster est en cours d'exécution, l'opération est plus complexe. Dans cette section, nous fournissons un exemple de programme Python qui montre comment installer ces bibliothèques.

Les exemples d'action d'amorçage et de programme Python illustrés dans cette section utilisent un script bash enregistré dans Amazon S3 pour installer les bibliothèques sur tous les nœuds.

Le script référencé dans l'exemple suivant utilise `pip` pour installer `paramiko`, `nltk`, `scipy`, `scikit-learn` et `pandas` pour le noyau Python 3 :

```
#!/bin/bash

sudo python3 -m pip install boto3 paramiko nltk scipy scikit-learn pandas
```

Après avoir créé le script, chargez-le dans un emplacement dans Amazon S3 par exemple, `s3://mybucket/install-my-jupyter-libraries.sh`. Pour plus d'informations, consultez la section [Chargement d'objets](#) dans le Guide de l'utilisateur Amazon Simple Storage Service afin de pouvoir les utiliser dans votre action de démarrage ou dans votre programme Python.

Pour spécifier une action d'amorçage qui installe des bibliothèques sur tous les nœuds lorsque vous créez un cluster à l'aide du AWS CLI

1. Créez un script semblable à l'exemple précédent et enregistrez-le dans un emplacement dans Amazon S3. Nous utilisons l'exemple `s3://mybucket/install-my-jupyter-libraries.sh`.
2. Créez le cluster avec JupyterHub et utilisez l'argument de l'option `--bootstrap-action` pour spécifier l'emplacement du script, comme indiqué dans l'exemple suivant :

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

```
aws emr create-cluster --name="MyJupyterHubCluster" --release-label emr-5.36.2 \  
--applications Name=JupyterHub --log-uri s3://MyBucket/MyJupyterClusterLogs \  
--use-default-roles --instance-type m5.xlarge --instance-count 2 --ec2-attributes \  
KeyName=MyKeyPair \  
--bootstrap-actions Path=s3://mybucket/install-my-jupyter- \  
libraries.sh,Name=InstallJupyterLibs
```

Pour spécifier une action d'amorçage qui installe des bibliothèques sur tous les nœuds lorsque vous créez un cluster à l'aide de la console

1. Accédez à la nouvelle EMR console Amazon et sélectionnez **Basculer vers l'ancienne console** dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).
2. Choisissez **Créer un cluster** et **Go to advanced options** (Aller aux options avancées).
3. Spécifiez les paramètres pour **Software and Steps** (Logiciel et étapes) et **Hardware** (Matériel) selon les besoins de votre application.
4. Sur l'écran **General Cluster Settings** (Paramètres généraux de cluster), développez **Bootstrap Actions** (Actions d'amorçage).
5. Pour **Add bootstrap action** (Ajouter une action d'amorçage), sélectionnez **Custom action** (Action personnalisée), **Configure and add** (Configurer et ajouter).
6. Pour **Name** (Nom), saisissez un nom convivial. Pour **Emplacement du script**, entrez l'emplacement de votre script dans Amazon S3 (l'exemple que nous utilisons est `s3://mybucket/install-my-jupyter-libraries.sh`). Laissez les **Optional arguments** (Arguments facultatifs) vides et choisissez **Add** (Ajouter).
7. Spécifiez les autres paramètres de votre cluster, puis choisissez **Next** (Suivant).
8. Spécifiez les paramètres de sécurité, puis choisissez **Create cluster** (Créer le cluster).

Exemple Installation de bibliothèques sur les nœuds principaux d'un cluster en cours d'exécution

Après avoir installé des bibliothèques sur le nœud principal à partir de Jupyter, vous pouvez installer des bibliothèques sur les nœuds principaux en cours d'exécution de diverses manières. L'exemple suivant présente un programme Python écrit pour s'exécuter sur une machine locale. Lorsque vous exécutez le programme Python localement, il utilise le symbole `AWS-RunShellScript` de AWS Systems Manager pour exécuter l'exemple de script, présenté plus haut dans cette section, qui installe des bibliothèques sur les nœuds principaux du cluster.

```
import argparse
import time
import boto3

def install_libraries_on_core_nodes(cluster_id, script_path, emr_client, ssm_client):
    """
    Copies and runs a shell script on the core nodes in the cluster.

    :param cluster_id: The ID of the cluster.
    :param script_path: The path to the script, typically an Amazon S3 object URL.
    :param emr_client: The Boto3 Amazon EMR client.
    :param ssm_client: The Boto3 AWS Systems Manager client.
    """
    core_nodes = emr_client.list_instances(
        ClusterId=cluster_id, InstanceGroupTypes=["CORE"]
    )["Instances"]
    core_instance_ids = [node["Ec2InstanceId"] for node in core_nodes]
    print(f"Found core instances: {core_instance_ids}.")

    commands = [
        # Copy the shell script from Amazon S3 to each node instance.
        f"aws s3 cp {script_path} /home/hadoop",
        # Run the shell script to install libraries on each node instance.
        "bash /home/hadoop/install_libraries.sh",
    ]
    for command in commands:
        print(f"Sending '{command}' to core instances...")
        command_id = ssm_client.send_command(
            InstanceIds=core_instance_ids,
            DocumentName="AWS-RunShellScript",
            Parameters={"commands": [command]},
            TimeoutSeconds=3600,
        )["Command"]["CommandId"]
```

```
while True:
    # Verify the previous step succeeded before running the next step.
    cmd_result = ssm_client.list_commands(CommandId=command_id)["Commands"][0]
    if cmd_result["StatusDetails"] == "Success":
        print(f"Command succeeded.")
        break
    elif cmd_result["StatusDetails"] in ["Pending", "InProgress"]:
        print(f"Command status is {cmd_result['StatusDetails']}, waiting...")
        time.sleep(10)
    else:
        print(f"Command status is {cmd_result['StatusDetails']}, quitting.")
        raise RuntimeError(
            f"Command {command} failed to run. "
            f"Details: {cmd_result['StatusDetails']}"
        )

def main():
    parser = argparse.ArgumentParser()
    parser.add_argument("cluster_id", help="The ID of the cluster.")
    parser.add_argument("script_path", help="The path to the script in Amazon S3.")
    args = parser.parse_args()

    emr_client = boto3.client("emr")
    ssm_client = boto3.client("ssm")

    install_libraries_on_core_nodes(
        args.cluster_id, args.script_path, emr_client, ssm_client
    )

if __name__ == "__main__":
    main()
```

JupyterHub historique des versions

Le tableau suivant répertorie la version JupyterHub incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

JupyterHub informations sur la version

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-7.2.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.36.2	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-7.1.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
		yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-7.0.0	1.5.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.15.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.14.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.13.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.12.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.11.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.11.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.10.1	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.10.0	1.5.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.9.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.9.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.8.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.8.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.7.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.36.1	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.36.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.6.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.35.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.5.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.4.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.3.1	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.3.0	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.2.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.2.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.1.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.1.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-6.0.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-6.0.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.34.0	1.4.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.33.1	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.33.0	1.2.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.32.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.32.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.31.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.31.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.30.2	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.30.1	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.30.0	1.1.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
mr-5.29.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.28.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.28.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.27.1	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.27.0	1.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.26.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.25.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.24.1	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.24.0	0.9.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.23.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.23.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.22.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.21.2	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.21.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.21.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.20.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.20.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.19.1	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.19.0	0.9.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.18.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.18.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.17.2	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.17.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.17.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.16.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.16.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.15.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.15.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.14.2	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub
emr-5.14.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Étiquette Amazon EMR Release	JupyterHub Version	Composants installés avec JupyterHub
emr-5.14.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

Apache Livy

Livy permet l'interaction via une REST interface avec un EMR cluster exécutant Spark. Vous pouvez utiliser l'interface REST ou une bibliothèque RPC cliente pour soumettre des tâches Spark ou des extraits de code Spark, récupérer les résultats de manière synchrone ou asynchrone et gérer Spark Context. Pour plus d'informations, consultez le [site web Apache Livy](#). Livy est inclus dans les EMR versions 5.9.0 et ultérieures d'Amazon.

Pour accéder à l'interface Web de Livy, configurez un SSH tunnel vers le nœud principal et une connexion proxy. Pour plus d'informations, voir [Afficher les interfaces Web hébergées sur EMR des clusters](#).

Le tableau suivant répertorie la version de Livy incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Livy.

Pour la version des composants installés avec Livy dans cette version, voir Versions des [composants de la version 7.2.0](#).

Informations sur la version de Livy pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Livy	Composants installés avec Livy
emr-7.2.0	Livy 0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Le tableau suivant répertorie la version de Livy incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Livy.

Pour connaître la version des composants installés avec Livy dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Livy pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Livy	Composants installés avec Livy
emr-6.15.0	Livy 0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Le tableau suivant répertorie la version de Livy incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Livy.

Pour la version des composants installés avec Livy dans cette version, voir la [version 5.36.2](#) Versions des composants.

Informations sur la version de Livy pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Livy	Composants installés avec Livy
emr-5.36.2	Livy 0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-

Étiquette EMR de sortie Amazon	Version Livy	Composants installés avec Livy
		yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Rubriques

- [Activation HTTPS avec Apache Livy](#)
- [Historique des versions de Livy](#)

Activation HTTPS avec Apache Livy

1. Provisionnez un EMR cluster Amazon avec le chiffrement du transit activé. Pour en savoir plus sur le chiffrement, consultez [Chiffrer les données au repos et en transit](#).
2. Créez un fichier nommé `livy_ssl.sh` avec le contenu suivant.

```
#!/bin/bash

KEystore_FILE=`awk '/ssl.server.keystore.location/{getline; print}' /etc/hadoop/conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`
KEystore_PASS=`awk '/ssl.server.keystore.password/{getline; print}' /etc/hadoop/conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`
KEY_PASS=`awk '/ssl.server.keystore.keypassword/{getline; print}' /etc/hadoop/conf/ssl-server.xml | sed -e 's/<[^>]*>//g' | tr -d ' \t\n\r\f'`

echo "livy.keystore $KEystore_FILE
livy.keystore.password $KEystore_PASS
livy.key-password $KEY_PASS" | sudo tee -a /etc/livy/conf/livy.conf >/dev/null

sudo systemctl restart livy-server.service
```

3. Exécutez le script suivant en tant qu'EMRÉtape Amazon. Ce script est modifié `/etc/livy/conf/livy.conf` pour être activéSSL.

```
--steps '[{"Args":["s3://DOC-EXAMPLE-BUCKET/
livy_ssl.sh"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"s3://
us-east-1.elasticmapreduce/libs/script-runner/script-
runner.jar","Properties":"","Name":"Custom JAR"}]'
```

4. Redémarrez le service Apache Livy afin que la modification prenne effet. Pour redémarrer Apache Livy, consultez la section [Arrêt et redémarrage des processus](#).
5. Testez que les clients peuvent désormais communiquer en utilisantHTTPS. Pour soumettre une tâche, par exemple, exécutez le code suivant.

```
curl -k -X POST --data '{"file": "local:///usr/lib/spark/examples/jars/spark-
examples.jar",
"className": "org.apache.spark.examples.SparkPi"}' \
-H "Content-Type: application/json" \
https://EMR_Master_Node_Host:8998/batches
```

Si vous l'avez activée HTTPS avec succès, Livy envoie une réponse indiquant que la commande a été acceptée et que le traitement par lots a été soumis.

```
{"id":1,"name":null,"owner":null,"proxyUser":null,"state":"starting","appId":null,"appInfo"
{"driverLogUrl":null,"sparkUiUrl":null},"log":["stdout: ", "\nstderr: ", "\nYARN
Diagnostics: "]}
```

Historique des versions de Livy

Le tableau suivant répertorie la version de Livy incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Livy

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-7.2.0	0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.36.2	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-7.1.0	0.8.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
		yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-7.0.0	0.7.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.15.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.14.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.13.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.12.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.11.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.11.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.10.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.10.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.9.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.9.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.8.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.8.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.7.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.36.1	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.36.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.6.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.35.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.5.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.4.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.3.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.3.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.2.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.2.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.1.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.1.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-6.0.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-6.0.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.34.0	0.7.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.33.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.33.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.32.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.32.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.31.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.31.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.30.2	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.30.1	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.30.0	0.7.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
mr-5.29.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.28.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.28.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.27.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.27.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.26.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.25.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.24.1	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.24.0	0.6.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.23.1	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.23.0	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.22.0	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.21.2	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.21.1	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.21.0	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.20.1	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.20.0	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.19.1	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.19.0	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.18.1	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.18.0	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, nginx
emr-5.17.2	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.17.1	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.17.0	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.16.1	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.16.0	0,5,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.15.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.15.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.14.2	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.14.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.14.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.13.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.13.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.12.3	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.12.2	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.12.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.12.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.4	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.11.3	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.2	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.11.1	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.11.0	0,4,0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.10.1	0,4,0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.10.0	0,4,0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Étiquette Amazon EMR Release	Version Livy	Composants installés avec Livy
emr-5.9.1	0,4,0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server
emr-5.9.0	0,4,0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server

Apache MXNet

Apache MXNet est une bibliothèque d'accélération conçue pour créer des réseaux de neurones et d'autres applications d'apprentissage profond. MXNet automatise les flux de travail courants et optimise les calculs numériques. MXNet vous aide à concevoir des architectures de réseaux neuronaux sans avoir à vous concentrer sur la mise en œuvre de calculs de bas niveau, tels que les opérations d'algèbre linéaire. MXNet est inclus dans les EMR versions 5.10.0 et ultérieures d'Amazon.

Pour plus d'informations, consultez le [site MXNet Web d'Apache](#).

Le tableau suivant répertorie la version MXNet incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants avec lesquels Amazon EMR installe. MXNet

Pour la version des composants installés MXNet dans cette version, voir [Versions des composants de la version 7.2.0](#).

MXNet informations de version pour emr-7.2.0

Étiquette EMR de sortie Amazon	MXNetVersion	Composants installés avec MXNet
emr-7.2.0	MXNet 1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Le tableau suivant répertorie la version MXNet incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants avec lesquels Amazon EMR installe. MXNet

Pour connaître la version des composants installés MXNet dans cette version, consultez la [version 6.15.0 Versions des composants](#).

MXNet informations de version pour emr-6.15.0

Étiquette EMR de sortie Amazon	MXNetVersion	Composants installés avec MXNet
emr-6.15.0	MXNet 1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Le tableau suivant répertorie la version MXNet incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants avec lesquels Amazon EMR installe. MXNet

Pour la version des composants installés MXNet dans cette version, voir la [version 5.36.2 Versions des composants](#).

MXNet informations de version pour emr-5.36.2

Étiquette EMR de sortie Amazon	MXNetVersion	Composants installés avec MXNet
emr-5.36.2	MXNet 1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

MXNet historique des versions

Le tableau suivant répertorie la version MXNet incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

MXNet informations sur la version

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-7.2.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.36.2	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-7.1.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
		server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-7.0.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.15.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.14.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-6.13.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.12.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.11.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-6.11.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.10.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.10.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-6.9.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.9.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.8.1	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-6.8.0	1.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.7.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.36.1	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.36.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.6.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.35.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-6.5.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.4.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.3.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-6.3.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.2.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.2.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-6.1.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.1.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-6.0.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-6.0.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.34.0	1.8.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.33.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.33.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.32.1	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.32.0	1.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.31.1	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.31.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.30.2	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.30.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.30.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
mr-5.29.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.28.1	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.28.0	1.5.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.27.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.27.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.26.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.25.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.24.1	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.24.0	1.4.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.23.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.23.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.22.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.21.2	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.21.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.21.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.20.1	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.20.0	1.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.19.1	1.3.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.19.0	1.3.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.18.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.18.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.17.2	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.17.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.17.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.16.1	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.16.0	1.2.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.15.1	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.15.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.14.2	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.14.1	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.14.0	1.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.13.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.13.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.3	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.12.2	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.1	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.12.0	1.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.11.4	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet
emr-5.11.3	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet
emr-5.11.2	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.11.1	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.11.0	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet
emr-5.10.1	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Étiquette Amazon EMR Release	MXNetVersion	Composants installés avec MXNet
emr-5.10.0	0.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Apache Oozie

Utilisez le planificateur de flux de travail Apache Oozie pour gérer et coordonner des travaux Hadoop. Pour plus d'informations, consultez <http://oozie.apache.org/>.

L'interface Web native d'Oozie n'est pas prise en charge sur Amazon EMR. Pour utiliser une interface frontale pour Oozie, essayez l'application Hue Oozie. Pour plus d'informations, consultez [Hue](#). Oozie est inclus dans les EMR versions 5.0.0 et ultérieures d'Amazon. Oozie est inclus sous la forme d'une application sandbox dans les versions précédentes. Pour plus d'informations, consultez [Versions publiées par Amazon EMR 4.x](#).

Si vous utilisez un Amazon Linux personnalisé AMI basé sur un Amazon Linux AMI dont la date de création est le 11/08/2018, le serveur Oozie ne démarre pas. Si vous utilisez Oozie, créez une personnalisation AMI basée sur un AMI identifiant Amazon Linux avec une date de création différente. Vous pouvez utiliser la AWS CLI commande suivante pour renvoyer une liste d'images IDs pour tous les HVM Amazon Linux AMIs avec une version 2018.03, ainsi que la date de sortie, afin de pouvoir choisir un Amazon Linux approprié AMI comme base. MyRegion Remplacez-le par l'identifiant de votre région, tel que us-west-2.

```
aws ec2 --region MyRegion describe-images --owner amazon --query 'Images[?
Name!=`null`][?starts_with(Name, `amzn-ami-hvm-2018.03`) == `true`].
[CreationDate,ImageId,Name]' --output text | sort -rk1
```

Le tableau suivant répertorie la version d'Oozie incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Oozie.

Pour la version des composants installés avec Oozie dans cette version, consultez la version [7.2.0 des versions](#) des composants.

Informations sur la version d'Oozie pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Oozie	Composants installés avec Oozie
emr-7.2.0	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode,

Étiquette EMR de sortie Amazon	Version Oozie	Composants installés avec Oozie
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Le tableau suivant répertorie la version d'Oozie incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Oozie.

Pour connaître la version des composants installés avec Oozie dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version d'Oozie pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Oozie	Composants installés avec Oozie
emr-6.15.0	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Le tableau suivant répertorie la version d'Oozie incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Oozie.

Pour la version des composants installés avec Oozie dans cette version, consultez la [version 5.36.2](#) Versions des composants.

Informations sur la version d'Oozie pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Oozie	Composants installés avec Oozie
emr-5.36.2	Oozie 5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Rubriques

- [Utiliser Oozie avec une base de données distante sur Amazon RDS](#)
- [Configurer la version Java pour Oozie](#)
- [Historique des versions d'Oozie](#)

Utiliser Oozie avec une base de données distante sur Amazon RDS

Par défaut, les informations utilisateur et l'historique des requêtes d'Oozie sont stockés dans une SQL base de données My locale sur le nœud principal. Vous pouvez également créer un ou plusieurs clusters compatibles Oozie à l'aide d'une configuration stockée dans Amazon S3 et d'une base de données My database dans Amazon Relational Database Service (Amazon). RDS Cela vous permet de conserver les informations utilisateur et l'historique des requêtes créés par Oozie sans

faire fonctionner votre EMR cluster Amazon. Nous vous recommandons d'utiliser le chiffrement côté serveur Amazon S3 pour stocker le fichier de configuration.

Commencez par créer la base de données distante pour Oozie.

Pour créer la base de données externe Ma SQL base de données

1. Ouvrez la RDS console Amazon à l'adresse <https://console.aws.amazon.com/rds/>.
2. Sélectionnez Launch a DB Instance (Lancer une instance de base de données).
3. Choisissez Mon, SQL puis sélectionnez Sélectionner.
4. Conservez la sélection par défaut de déploiement multi-AZ et de IOPS stockage provisionné et choisissez Next.
5. Conservez les valeurs par défaut des spécifications de l'instance, spécifiez les paramètres, puis choisissez Next (Suivant).
6. Dans la page de configuration de paramètres avancés, choisissez des noms de groupe de sécurité et de base de données adéquats. Le groupe de sécurité que vous utilisez doit au moins autoriser l'accès entrant au port 3306 depuis le nœud principal de votre cluster. Si vous n'avez pas créé votre cluster à ce stade, vous pouvez autoriser tous les hôtes à se connecter au port 3306 et ajuster le groupe de sécurité, une fois que vous avez lancé le cluster. Choisissez Launch DB Instance (Lancer l'instance de base de données).
7. Dans le RDS tableau de bord, sélectionnez Instances et sélectionnez l'instance que vous venez de créer. Lorsque votre base de données est disponible, notez le nom de base de données, le nom d'utilisateur, le mot de passe et le nom d'hôte de l'instance RDS. Vous allez utiliser ces informations pour créer et configurer votre cluster.

Pour spécifier une base de SQL données My Database externe pour Oozie lors du lancement d'un cluster à l'aide du AWS CLI

Pour spécifier une base de SQL données My Database externe pour Oozie lors du lancement d'un cluster à l'aide du AWS CLI, utilisez les informations que vous avez notées lors de la création de votre RDS instance pour la configurer `oozie-site` avec un objet de configuration.

Note

Vous pouvez créer plusieurs clusters qui utilisent la même base de données externe, mais chaque cluster partage l'historique des requêtes et les informations utilisateur.

- À l'aide de AWS CLI, créez un cluster sur lequel Oozie est installé, en utilisant la base de données externe que vous avez créée et en référençant un fichier de configuration avec une classification de configuration pour Oozie qui spécifie les propriétés de la base de données. L'exemple suivant crée un cluster avec Oozie installé en référençant un fichier de configuration dans Amazon S3, `myConfig.json`, qui spécifie la configuration de la base de données.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Oozie
Name=Spark Name=Hive \
--instance-type m5.xlarge --instance-count 3 \
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-
default-roles
```

Vous trouverez ci-dessous des exemples de contenus du fichier `myConfig.json`. Remplacez *JDBC URL*, *username*, et *password* avec le nom JDBC URL d'utilisateur et le mot de passe de votre RDS instance.

Important

Ils JDBC URL doivent inclure le nom de la base de données en tant que suffixe. Par exemple, `jdbc:mysql://.xxxxxxxxx.us-east-1.rds.amazonaws.com:3306/dbname.oozie-external-db`

```
[{
  "Classification": "oozie-site",
  "Properties": {
    "oozie.service.JPAService.jdbc.driver": "org.mariadb.jdbc.Driver",
    "oozie.service.JPAService.jdbc.url": "JDBC URL",

    "oozie.service.JPAService.jdbc.username": "username",
    "oozie.service.JPAService.jdbc.password": "password"
  }
}]
```

```
    },  
    "Configurations": []  
  ]  
}
```

Configurer la version Java pour Oozie

Oozie exécute plusieurs processus Java Virtual Machine (JVM). Cette page explique comment configurer la version Java pour chaque processus.

- **Serveur Oozie** : définissez `JAVA_HOME` dans la classification `oozie-env` pour mettre à jour la version Java pour `EmbeddedOozieServer`.
- **Oozie Launcher AM** : Oozie Launcher AM est une tâche MR à mappeur unique qui invoque les bibliothèques clientes d'applications appropriées telles que Hadoop et Hive. Sauf configuration contraire, les versions d'exécution d'Oozie Launcher AM sont les mêmes que celles de Java pour Hadoop dans le cluster. EMR Pour configurer l'environnement d'exécution Java pour Oozie Launcher AM, définissez la propriété suivante dans `workflow.xml` de la tâche :

```
<property>  
  <name>mapred.child.env</name>  
  <value>JAVA_HOME=/path/to/JAVA_HOME</value>  
</property>
```

Cette propriété garantit que le lanceur Oozie AM pour le job Oozie s'exécute sur la version Java que vous spécifiez, plutôt que sur la version Java définie dans Hadoop.

- **Exécutable du client d'application** : étant donné qu'Oozie Launcher AM appelle le client de l'application par défaut, le runtime Java de l'exécutable client est le même que celui d'Oozie Launcher AM.
- **Applications lancées par une tâche Oozie** : sauf indication contraire, les versions d'exécution de l'application elle-même JVMs qui sont lancées par une tâche Oozie sont les mêmes que celles de Java pour Hadoop dans le cluster. EMR Selon le type d'action Oozie Workflow utilisé pour lancer l'application dans un job Oozie (action Spark ou Hive), vous pouvez mettre à jour le runtime Java par défaut pour l'application réelle JVMs dans le runtime Java par défaut pour le job Oozie. `workflow.xml`

Historique des versions d'Oozie

Le tableau suivant répertorie la version d'Oozie incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section [Version des composants de votre version](#) dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Oozie

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-7.2.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-5.36.2	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
		timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-7.1.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-7.0.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.15.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.14.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.13.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.12.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.11.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.11.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.10.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker
emr-6.10.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.9.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.9.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.8.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.8.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.7.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.36.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.36.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.6.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.35.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.5.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.4.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.3.1	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.3.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.2.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.2.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.1.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.1.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-6.0.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-6.0.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.34.0	5.2.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.33.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.33.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.32.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.32.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.31.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.31.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.30.2	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.30.1	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.30.0	5.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
mr-5.29.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.28.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.28.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.27.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.27.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.26.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.25.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.24.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.24.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.23.1	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.23.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.22.0	5.1.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.21.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.21.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.20.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.20.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.19.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.19.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.18.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.18.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.17.2	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.17.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.17.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.16.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.16.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.15.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.15.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.14.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.14.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.13.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.13.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.12.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.12.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.12.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.11.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.11.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.11.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.11.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.10.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.10.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.9.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.9.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.8.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.8.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.8.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.7.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.7.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.6.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.6.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.4	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.5.3	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.5.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.5.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.4.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.4.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.3.2	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.3.1	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.3.0	4.3.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.2.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.2.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.2.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.1.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.1.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.0.3	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.0.2	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn
emr-5.0.1	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Oozie	Composants installés avec Oozie
emr-5.0.0	4.2.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Notes de mise à jour d'Oozie par version

Rubriques

- [Amazon EMR 7.1.0 - Modifications apportées à Oozie](#)

Amazon EMR 7.1.0 - Modifications apportées à Oozie

Type	Description
Upgrade	Met à jour Netty vers la version 4.1.100.Final.
Upgrade	Met à jour Jetty vers la version 9.4.53.v20231009.

Apache Phoenix

Apache Phoenix est utilisé à OLTP des fins d'analyse opérationnelle, ce qui vous permet d'utiliser SQL des requêtes standard et JDBC APIs de travailler avec un magasin de HBase sauvegarde Apache. Pour plus d'informations, consultez [Phoenix en 15 minutes ou moins](#). Phoenix est inclus dans les EMR versions 4.7.0 et ultérieures d'Amazon.

Si vous effectuez une mise à niveau d'une version antérieure d'Amazon EMR vers la EMR version 5.4.0 ou ultérieure d'Amazon et que vous utilisez l'indexation secondaire, mettez à niveau les index locaux comme décrit dans la documentation d'[Apache Phoenix](#). Amazon EMR supprime les configurations requises de la `hbase-site classification`, mais les index doivent être remplis. La mise à niveau en ligne et hors ligne des index est prise en charge. Les mises à niveau sont par défaut en ligne, ce qui signifie que les index sont remplis de nouveau lors de l'initialisation depuis les clients Phoenix en version 4.8.0 ou supérieure. Pour spécifier des mises à niveau hors ligne, définissez la `phoenix.client.localIndexUpgrade` configuration sur `false` dans la `phoenix-site classification`, puis SSH sur le nœud principal à exécuter `psql [zookeeper] -1`.

Le tableau suivant répertorie la version de Phoenix incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Phoenix.

Pour la version des composants installés avec Phoenix dans cette version, voir [Versions des composants de la version 7.2.0](#).

Informations sur la version de Phoenix pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Phoenix	Composants installés avec Phoenix
emr-7.2.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Étiquette EMR de sortie Amazon	Version Phoenix	Composants installés avec Phoenix
		timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Le tableau suivant répertorie la version de Phoenix incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Phoenix.

Pour connaître la version des composants installés avec Phoenix dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Phoenix pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Phoenix	Composants installés avec Phoenix
emr-6.15.0	Phoenix 5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-

Étiquette EMR de sortie Amazon	Version Phoenix	Composants installés avec Phoenix
		server, zookeeper-client, zookeeper-server

Le tableau suivant répertorie la version de Phoenix incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Phoenix.

Pour la version des composants installés avec Phoenix dans cette version, voir la [version 5.36.2 Versions des composants](#).

Informations sur la version de Phoenix pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Phoenix	Composants installés avec Phoenix
emr-5.36.2	Phoenix 4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Rubriques

- [Création d'un cluster avec Phoenix](#)
- [Clients Phoenix](#)

- [Historique des sorties de Phoenix](#)

Création d'un cluster avec Phoenix

Vous installez Phoenix en choisissant l'application lors de la création d'un cluster sur la console ou en utilisant l' AWS CLI. Les procédures et exemples suivants montrent comment créer un cluster avec Phoenix et HBase. Pour plus d'informations sur la création de clusters à l'aide de la console, notamment sur les options avancées, consultez [Planifier et configurer des clusters](#) dans le guide EMR de gestion Amazon.

Pour lancer un cluster avec Phoenix installé à l'aide des Quick Options (Options rapides) pour la création d'un cluster sur la console

1. Ouvrez la EMR console Amazon à l'[adresse https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr).
2. Choisissez Create cluster (Créer un cluster) pour utiliser la Quick Create (Création rapide).
3. À côté du champ Configuration des logiciels, choisissez la version la plus récente appropriée à votre application. Phoenix apparaît en tant qu'option uniquement lorsque la Version emr-4.7.0 d'Amazon ou ultérieure est sélectionnée.
4. Pour les applications, choisissez la deuxième option HBase : HBase **ver** avec Ganglia **ver**, Hadoop **ver**, Ruche **ver**, Hué **ver**, Phénix **ver**, et ZooKeeper **ver**.
5. Sélectionnez d'autres options si nécessaire, puis choisissez Create cluster (Créer le cluster).

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

L'exemple suivant lance un cluster avec Phoenix installé à l'aide des paramètres de configuration par défaut.

Pour lancer un cluster avec Phoenix et HBase à l'aide du AWS CLI

- Créez le cluster à l'aide de la commande suivante :


```
aws emr create-cluster --name "Cluster with Phoenix" --release-label emr-7.2.0 \  
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Personnalisation des configurations Phoenix

Lors de la création d'un cluster, vous configurez Phoenix en définissant des valeurs `hbase-site.xml` à l'aide de la classification de configuration `hbase-site`.

Pour plus d'informations, consultez [Configuration et réglage](#) dans la documentation Phoenix.

L'exemple suivant montre comment utiliser un JSON fichier stocké dans Amazon S3 pour spécifier la valeur de `false` pour la `phoenix.schema.dropMetaData` propriété. Plusieurs propriétés peuvent être spécifiées pour une classification unique. Pour plus d'informations, consultez [Configuration des applications](#). La `create-cluster` commande fait ensuite référence au JSON fichier en tant que `--configurations` paramètre.

Le contenu du JSON fichier enregistré dans `/mybucket/myfolder/myconfig.json` est le suivant.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "phoenix.schema.dropMetaData": "false"  
    }  
  }  
]
```

La `create cluster` commande qui fait référence au JSON fichier est illustrée dans l'exemple suivant.

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Phoenix \  
Name=HBase --instance-type m5.xlarge --instance-count 2 \  
--configurations https://s3.amazonaws.com/mybucket/myfolder/myconfig.json
```

Note

Les demandes de reconfiguration pour toutes les classifications de configuration Phoenix ne sont prises en charge que dans les EMR versions 5.23.0 et ultérieures d'Amazon, et

non dans les EMR versions 5.21.0 ou 5.22.0 d'Amazon. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#)

Clients Phoenix

Vous vous connectez à Phoenix soit à l'aide d'un JDBC client construit avec des dépendances complètes, soit à l'aide du « client léger » qui utilise le serveur de requêtes Phoenix et ne peut être exécuté que sur un nœud maître d'un cluster (par exemple en utilisant un SQL client, une étape, une ligne de commande, une redirection de SSH port, etc.). Lorsque vous utilisez le JDBC client « fat », il doit toujours avoir accès à tous les nœuds du cluster car il se connecte directement aux HBase services. Le client Phoenix « léger » a uniquement besoin d'un accès au serveur de requête Phoenix dans un port par défaut 8765. Il existe plusieurs [scripts](#) au sein de Phoenix qui utilisent ces clients.

Utilisez une EMR étape Amazon pour effectuer une requête à l'aide de Phoenix

La procédure suivante permet de restaurer un instantané à partir de ces données HBase et de les utiliser pour exécuter une requête Phoenix. Vous pouvez étendre cet exemple ou créer un script qui s'appuie sur les clients Phoenix pour répondre à vos besoins.

1. Créez un cluster avec Phoenix installé, à l'aide de la commande suivante :

```
aws emr create-cluster --name "Cluster with Phoenix" --log-uri s3://myBucket/myLogFolder --release-label emr-7.2.0 \  
--applications Name=Phoenix Name=HBase --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

2. Créez, puis chargez les fichiers suivants sur Amazon S3 :

copySnapshot.sh

```
sudo su hbase -s /bin/sh -c 'hbase snapshot export \  
-D hbase.rootdir=s3://us-east-1.elasticmapreduce.samples/hbase-demo-customer-data/snapshot/ \  
-snapshot customer_snapshot1 \  
-copy-to hdfs://masterDNSName:8020/user/hbase \  
-mappers 2 -chuser hbase -chmod 700'
```

runQuery.sh

```
aws s3 cp s3://myBucket/phoenixQuery.sql /home/hadoop/  
/usr/lib/phenix/bin/sqlline-thin.py http://localhost:8765 /home/hadoop/  
phoenixQuery.sql
```

phoenixQuery.sql

Note

Vous ne devez inclure `COLUMN_ENCODED_BYTES=0` dans l'exemple suivant que si vous utilisez les EMR versions 5.26.0 ou supérieures d'Amazon.

```
CREATE VIEW "customer" (  
pk VARCHAR PRIMARY KEY,  
"address"."state" VARCHAR,  
"address"."street" VARCHAR,  
"address"."city" VARCHAR,  
"address"."zip" VARCHAR,  
"cc"."number" VARCHAR,  
"cc"."expire" VARCHAR,  
"cc"."type" VARCHAR,  
"contact"."phone" VARCHAR)  
COLUMN_ENCODED_BYTES=0;  
  
CREATE INDEX my_index ON "customer" ("customer"."state") INCLUDE("PK",  
"customer"."city", "customer"."expire", "customer"."type");  
  
SELECT "customer"."type" AS credit_card_type, count(*) AS num_customers FROM  
"customer" WHERE "customer"."state" = 'CA' GROUP BY "customer"."type";
```

Utilisez le AWS CLI pour envoyer les fichiers au compartiment S3 :

```
aws s3 cp copySnapshot.sh s3://myBucket/  
aws s3 cp runQuery.sh s3://myBucket/  
aws s3 cp phoenixQuery.sql s3://myBucket/
```

3. Créez une table à l'aide de l'étape suivante soumise au cluster que vous avez créé à l'étape 1 :

createTable.json

```
[
  {
    "Name": "Create HBase Table",
    "Args": ["bash", "-c", "echo '$'create \"customer\", \"address\", \"cc\", \"contact\" | hbase shell"],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./createTable.json
```

- Utilisez `script-runner.jar` pour exécuter le script `copySnapshot.sh` que vous avez téléchargé précédemment sur votre compartiment S3 :

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="HBase Copy Snapshot",ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/copySnapshot.sh"]
```

Cela exécute une MapReduce tâche pour copier vos données de capture d'écran dans le clusterHDFS.

- Restaurez l'instantané que vous avez copiée dans le cluster à l'aide de l'étape suivante :

restoreSnapshot.json

```
[
  {
    "Name": "restore",
    "Args": ["bash", "-c", "echo '$'disable \"customer\"; restore_snapshot \"customer_snapshot1\"; enable \"customer\" | hbase shell"],
    "Jar": "command-runner.jar",
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR"
  }
]
```

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps file:///./restoreSnapshot.json
```

6. Utilisez `script-runner.jar` pour exécuter le script `runQuery.sh` que vous avez téléchargé précédemment sur votre compartiment S3 :

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Phoenix Run Query",ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-
runner.jar,Args=["s3://myBucket/runQuery.sh"]
```

La requête s'exécute et renvoie les résultats au stdout de l'étape. La réalisation de cette étape peut prendre quelques minutes.

7. Vérifiez les résultats des étapes dans le journal URI que vous avez utilisé lors de la création du cluster à l'étape 1. stdout Les résultats doivent avoir l'aspect suivant :

```
+-----+-----+
|          CREDIT_CARD_TYPE          |          NUM_CUSTOMERS          |
+-----+-----+
| american_express                    | 5728                             |
| dankort                             | 5782                             |
| diners_club                         | 5795                             |
| discover                            | 5715                             |
| forbrugsforeningen                 | 5691                             |
| jcb                                 | 5762                             |
| laser                               | 5769                             |
| maestro                             | 5816                             |
| mastercard                         | 5697                             |
| solo                               | 5586                             |
| switch                             | 5781                             |
| visa                               | 5659                             |
+-----+-----+
```

Historique des sorties de Phoenix

Le tableau suivant répertorie la version de Phoenix incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section [Version des composants de votre version](#) dans [Versions](#)

[publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations de version de Phoenix

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-7.2.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.36.2	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
		region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-7.1.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-7.0.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.15.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.14.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.13.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.12.0	5.1.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.11.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.11.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.10.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.10.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.9.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.9.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.8.1	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.8.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.7.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.36.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.36.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.6.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.35.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.5.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.4.0	5.1.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.3.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.3.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.2.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.2.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.1.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.1.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-6.0.1	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-6.0.0	5.0.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.34.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.33.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.33.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.32.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.32.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.31.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.31.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.30.2	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.30.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.30.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
mr-5.29.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.28.1	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.28.0	4.14.3	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.27.1	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.27.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.26.0	4.14.2	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.25.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.24.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.24.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.23.1	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.23.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.22.0	4.14.1	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.21.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.21.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.21.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.20.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.20.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.19.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.19.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.18.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.18.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.17.2	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.17.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.17.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.16.1	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.16.0	4.14.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.15.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.15.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.14.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.14.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.13.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.13.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.12.3	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.2	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.12.1	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.12.0	4.13.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.11.4	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.11.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.11.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.11.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.10.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.10.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.9.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.9.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.3	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.8.2	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.8.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.8.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.7.1	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.7.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.6.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.6.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.4	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.5.3	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.2	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.5.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.5.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.4.1	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.4.0	4.9.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.3.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.3.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.3.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.2.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.2.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.2.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.1.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.1.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.0.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-5.0.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-5.0.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.6	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-4.9.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-4.9.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.9.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-4.9.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.5	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-4.8.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-4.8.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.8.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-4.8.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.4	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-4.7.3	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.2	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	Version Phoenix	Composants installés avec Phoenix
emr-4.7.1	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-4.7.0	4.7.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Apache Pig

Apache Pig est une bibliothèque Apache open source qui s'exécute au-dessus de Hadoop. Elle fournit un langage de script que vous pouvez utiliser pour transformer des ensembles de données volumineux, sans avoir à écrire de code complexe en langage informatique de niveau inférieur, comme Java. La bibliothèque prend SQL des commandes similaires écrites dans un langage appelé Pig Latin et les convertit en tâches Tez basées sur des graphes acycliques dirigés (DAGs) ou des MapReduce programmes. Pig fonctionne avec des données structurées et non structurées dans différents formats. Pour plus d'informations sur Pig, consultez la page <http://pig.apache.org/>.

Vous pouvez exécuter les commandes Pig de façon interactive ou en mode de traitement par lots. Pour utiliser Pig de manière interactive, créez une SSH connexion au nœud principal et soumettez des commandes à l'aide du shell Grunt. Pour utiliser Pig en mode de traitement par lots, écrivez vos scripts Pig, chargez-les sur Amazon S3 et envoyez-les en tant qu'étapes du cluster. Pour plus d'informations sur la soumission d'un travail à un cluster, consultez [Soumettre un travail à un cluster](#) dans le Guide EMR de gestion Amazon.

Lorsque vous utilisez Pig pour écrire une sortie dans une HCatalog table dans Amazon S3, désactivez Amazon EMR Direct Write en définissant les `mapred.output.direct.EmrFileSystem` propriétés `mapred.output.direct.NativeS3FileSystem` et `surfalse`. Pour plus d'informations, consultez [En utilisant HCatalog](#). Dans un script Pig, vous pouvez utiliser les commandes SET `mapred.output.direct.NativeS3FileSystem false` et SET `mapred.output.direct.EmrFileSystem false`.

Le tableau suivant répertorie la version de Pig incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Pig.

Pour la version des composants installés avec Pig dans cette version, voir [Versions des composants de la version 7.2.0](#).

Informations sur la version de Pig pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version de Pig	Composants installés avec Pig
emr-7.2.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-

Étiquette EMR de sortie Amazon	Version de Pig	Composants installés avec Pig
		cp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Le tableau suivant répertorie la version de Pig incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Pig.

Pour connaître la version des composants installés avec Pig dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Pig pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version de Pig	Composants installés avec Pig
emr-6.15.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Le tableau suivant répertorie la version de Pig incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Pig.

Pour la version des composants installés avec Pig dans cette version, voir la [version 5.36.2 Versions des composants](#).

Informations sur la version de Pig pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version de Pig	Composants installés avec Pig
emr-5.36.2	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Rubriques

- [Soumission de travail Pig](#)
- [Appel de fonctions définies par l'utilisateur à partir de Pig](#)
- [Historique de versions Pig](#)

Soumission de travail Pig

Cette section explique comment soumettre le travail de Pig à un EMR cluster Amazon. Les exemples suivants génèrent un rapport contenant le nombre total d'octets transférés, une liste des 50 adresses IP principales, une liste des 50 référents externes principaux et les 50 critères de recherche principaux sur Bing et Google. Le script Pig est situé dans le compartiment Amazon S3 `s3://elasticmapreduce/samples/pig-apache/do-reports2.pig`. Les données d'entrée

sont situées dans le compartiment Amazon S3 `s3://elasticmapreduce/samples/pig-apache/input`. La sortie est enregistrée dans un compartiment Amazon S3.

Soumettez le travail de Pig à l'aide de la EMR console Amazon

Cet exemple décrit comment utiliser la EMR console Amazon pour ajouter une étape Pig à un cluster.

Pour soumettre une étape Pig

1. Ouvrez la EMR console Amazon à l'[adresse https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr).
2. Choisissez Créer un cluster pour créer un cluster avec Pig installé. Pour savoir comment créer un cluster, consultez [Planifier et configurer un EMR cluster Amazon](#).
3. Ouvrez un terminal et accédez au nœud principal de votre cluster SSH en suivant les étapes décrites dans la section [Se connecter au nœud principal à l'aide de SSH](#). Une fois que vous avez fait cela, exécutez les étapes suivantes.

```
sudo mkdir -p /home/hadoop/lib/pig/  
sudo aws s3 cp s3://elasticmapreduce/libs/pig/0.3/piggybank-0.3-amzn.jar /home/  
hadoop/lib/pig/piggybank.jar
```

4. Dans la console, cliquez sur Liste des clusters et sélectionnez le nom du cluster que vous avez créé.
5. Faites défiler l'affichage jusqu'à la section Étapes et développez-la, puis choisissez Ajouter une étape.
6. Dans la boîte de dialogue Ajouter une étape :
 - Pour Type d'étape, choisissez Programme Pig.
 - Pour Nom, acceptez le nom par défaut (Programme Pig) ou saisissez un nouveau nom.
 - Pour Emplacement S3 du script, saisissez l'emplacement du script Pig. Par exemple : **s3://elasticmapreduce/samples/pig-apache/do-reports2.pig**.
 - Pour Emplacement S3 d'entrée, saisissez l'emplacement des données d'entrée. Par exemple : **s3://elasticmapreduce/samples/pig-apache/input**.
 - Pour Emplacement S3 de sortie, saisissez le nom du compartiment de sortie Amazon S3 ou naviguez à ce dernier.
 - Pour Arguments, laissez le champ vide.
 - Pour Action sur échec, acceptez l'option par défaut (Continuer).
7. Choisissez Ajouter. L'étape s'affiche dans la console avec le statut En suspens.

8. Le statut de l'étape passe de Pending (En suspens) à Running (En cours d'exécution) puis à Completed (Terminé), au fur et à mesure de son exécution. Pour mettre à jour l'état, choisissez l'icône d'Actualisation au-dessus de la colonne Actions. Lorsque votre étape est terminée, vérifiez votre compartiment Amazon S3 pour confirmer que les fichiers de sortie de votre étape Pig s'y trouvent.

Soumettez le travail de Pig à l'aide du AWS CLI

Pour soumettre une étape Pig à l'aide du AWS CLI

Lorsque vous lancez un cluster à l'aide du paramètre AWS CLI, utilisez le `--applications` paramètre pour installer Pig. Pour soumettre une étape Pig, utilisez le paramètre `--steps`.

1. Pour lancer un cluster sur lequel Pig est installé, tapez la commande suivante, en remplaçant *myKey* and *DOC-EXAMPLE-BUCKET/* avec le nom de votre paire de EC2 clés et de votre compartiment Amazon S3.

```
aws emr create-cluster \  
--name "Test cluster" \  
--log-uri s3://DOC-EXAMPLE-BUCKET/ \  
--release-label emr-5.36.2 \  
--applications Name=Pig \  
--use-default-roles \  
--ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge \  
--instance-count 3
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

Lorsque vous spécifiez le nombre d'instances sans utiliser le paramètre `--instance-groups`, un seul nœud maître est lancé et les instances restantes sont lancées en tant que nœuds principaux. Tous les nœuds utilisent le type d'instance spécifié dans la commande.

Note

Si vous n'avez pas encore créé le rôle de EMR service et le profil d'EC2instance par défaut, tapez `aws emr create-default-roles` pour les créer avant de taper la `create-cluster` sous-commande.

2. Pour soumettre une étape Pig, entrez la commande suivante, en remplaçant *myClusterId* and *DOC-EXAMPLE-BUCKET* avec votre ID de cluster et le nom de votre compartiment Amazon S3.

```
aws emr add-steps \  
--cluster-id myClusterId \  
--steps Type=PIG,Name="Pig Program",ActionOnFailure=CONTINUE,Args=[-f,s3://  
elasticmapreduce/samples/pig-apache/do-reports2.pig,-p,INPUT=s3://elasticmapreduce/  
samples/pig-apache/input,-p,OUTPUT=s3://DOC-EXAMPLE-BUCKET/pig-apache/output]
```

Cette commande renverra un identifiant d'étape, que vous pourrez utiliser pour vérifier le State de votre étape.

3. Vérifiez l'état de votre étape à l'aide de la commande `describe-step`.

```
aws emr describe-step --cluster-id myClusterId --step-id s-1XXXXXXXXXXA
```

La valeur State de l'étape passe de PENDING à RUNNING, puis à COMPLETED, au fur et à mesure que l'étape s'exécute. Lorsque votre étape est terminée, vérifiez votre compartiment Amazon S3 pour confirmer que les fichiers de sortie de votre étape Pig s'y trouvent.

Pour plus d'informations sur l'utilisation EMR des commandes Amazon dans le AWS CLI, consultez la [référence des AWS CLI commandes](#).

Appel de fonctions définies par l'utilisateur à partir de Pig

Pig permet d'appeler des fonctions définies par l'utilisateur (UDFs) depuis les scripts Pig. Vous pouvez le faire pour implémenter un traitement personnalisé à utiliser dans vos scripts Pig. Les langages actuellement pris en charge sont Java, Python/Jython et JavaScript (bien que le JavaScript support soit encore expérimental).

Les sections suivantes expliquent comment enregistrer vos fonctions avec Pig afin de pouvoir les appeler à partir du shell Pig ou de scripts Pig. Pour plus d'informations sur l'utilisation UDFs de Pig, consultez [la documentation de Pig](#) correspondant à votre version de Pig.

JARFichiers d'appels depuis Pig

Vous pouvez utiliser des JAR fichiers personnalisés avec Pig en utilisant la REGISTER commande de votre script Pig. Il s'agit d'un JAR fichier local ou d'un système de fichiers distant tel qu'Amazon S3. Lorsque le script Pig s'exécute, Amazon EMR télécharge le JAR fichier automatiquement sur le nœud principal, puis le télécharge dans le JAR cache distribué Hadoop. Ainsi, le JAR fichier est automatiquement utilisé selon les besoins par toutes les instances du cluster.

Pour utiliser des JAR fichiers avec Pig

1. Chargez votre JAR fichier personnalisé dans Amazon S3.
2. Utilisez la REGISTER commande de votre script Pig pour spécifier le compartiment du JAR fichier personnalisé sur Amazon S3.

```
REGISTER s3://mybucket/path/mycustomjar.jar;
```

Appel de scripts Python/Jython à partir de Pig

Vous pouvez enregistrer des scripts Python avec Pig, puis appeler des fonctions dans ces scripts à partir du shell Pig ou dans un script Pig. Pour cela, vous devez spécifier l'emplacement du script avec le mot clé `register`.

Etant donné que Pig est écrit en Java, il utilise le moteur de script Jython pour analyser les scripts Python. Pour plus d'informations sur Jython, accédez à <http://www.jython.org/>.

Pour appeler un script Python/Jython à partir de Pig

1. Écrivez un script Python et chargez-le dans un emplacement sur Amazon S3. Il doit s'agir d'un compartiment appartenant au même compte que celui qui crée le cluster Pig, ou qui dispose des autorisations définies pour que le compte qui a créé le cluster puisse y accéder. Dans cet exemple, le script est chargé sur `s3://mybucket/pig/python`.
2. Démarrez un cluster Pig. Si vous accédez à Pig à partir du shell Grunt, exécutez un cluster interactif. Si vous exécutez les commandes Pig à partir d'un script, démarrez un cluster Pig

scripté. Cet exemple démarre un cluster interactif. Pour plus d'informations sur la création d'un cluster Pig, consultez [Soumission de travail Pig](#).

3. Pour un cluster interactif, utilisez-le SSH pour vous connecter au nœud principal et exécuter le shell Grunt. Pour plus d'informations, reportez-vous à la section [SSH into the master node](#).
4. Exécutez le shell Grunt pour Pig en tapant `pig` dans la ligne de commande :

```
pig
```

5. Enregistrez la bibliothèque Jython et votre script Python avec Pig à l'aide du mot clé `register` à l'invite de commande Grunt, comme illustré dans la commande suivante, dans laquelle vous devez spécifier l'emplacement de votre script dans Amazon S3 :

```
grunt> register 'lib/jython.jar';  
grunt> register 's3://mybucket/pig/python/myscript.py' using jython as myfunctions;
```

6. Chargez les données d'entrée. L'exemple suivant charge l'entrée à partir d'un emplacement Amazon S3 :

```
grunt> input = load 's3://mybucket/input/data.txt' using TextLoader as  
(line:chararray);
```

7. Vous pouvez désormais appeler des fonctions dans votre script à partir de Pig en y faisant référence à l'aide de `myfunctions` :

```
grunt> output=foreach input generate myfunctions.myfunction($1);
```

Historique de versions Pig

Le tableau suivant répertorie la version de Pig incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Pig

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-7.2.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-5.36.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-7.1.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
		server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-7.0.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.15.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.14.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.13.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.12.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.11.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.11.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.10.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.10.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker
emr-6.9.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.9.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.8.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.8.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.7.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.36.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.36.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.6.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.35.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.5.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.4.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.3.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.3.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.2.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.2.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-6.1.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-6.1.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.34.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.33.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.33.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.32.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.32.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.31.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.31.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.30.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.30.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.30.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
mr-5.29.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.28.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.28.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.27.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.27.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.26.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.25.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.24.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.24.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.23.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.23.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.22.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.21.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.21.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.21.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.20.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.20.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.19.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.19.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.18.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.18.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.17.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.17.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.17.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.16.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.16.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.15.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.15.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.14.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.14.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.14.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.13.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.13.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.12.3	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.12.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.12.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.12.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.11.4	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.11.3	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.11.2	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.11.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.11.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.10.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.10.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.9.1	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.9.0	0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.8.3	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.8.2	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.8.1	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.8.0	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.7.1	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.7.0	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.6.1	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.6.0	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.5.4	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.5.3	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.5.2	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.5.1	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.5.0	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.4.1	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.4.0	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.3.2	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.3.1	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.3.0	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.2.3	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.2.2	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.2.1	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.2.0	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.1.1	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.1.0	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.0.3	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.0.2	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-5.0.1	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn
emr-5.0.0	0,16,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-4.9.6	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.9.5	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.9.4	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-4.9.3	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.9.2	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.9.1	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-4.8.5	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.8.4	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.8.3	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-4.8.2	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.8.1	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.8.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-4.7.4	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.7.3	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client
emr-4.7.2	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, pig-client

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-4.7.1	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.7.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.6.1	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-4.6.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.5.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.4.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-4.3.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.2.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client
emr-4.1.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httplibfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Étiquette Amazon EMR Release	Version de Pig	Composants installés avec Pig
emr-4.0.0	0,14,0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano-de, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Presto et Trino

Note

Presto SQL a été renommé Trino en décembre 2020. EMR Les versions 6.4.0 et ultérieures d'Amazon utilisent le nom Trino, tandis que les versions antérieures utilisent le nom Presto. SQL

[Presto](#) est un moteur de SQL requêtes rapide conçu pour les requêtes analytiques interactives sur de grands ensembles de données provenant de sources multiples. Pour plus d'informations, consultez le [site web Presto](#). Presto est inclus dans les EMR versions 5.0.0 et ultérieures d'Amazon. Les versions antérieures incluent Presto en tant qu'application d'environnement de test (sandbox). Pour de plus amples informations, veuillez consulter [Versions publiées par Amazon EMR 4.x](#). Les EMR versions 6.1.0 et ultérieures d'Amazon prennent en charge [Trino](#) (PrestoSQL) en plus de Presto. Pour de plus amples informations, veuillez consulter [Installation de PrestoDB et Trino](#).

Le tableau suivant répertorie la version de Presto incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Presto.

Pour la version des composants installés avec Presto dans cette version, voir Versions des [composants de la version 7.2.0](#).

Informations sur la version Presto pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Presto	Composants installés avec Presto
emr-7.2.0	Presto 0.285	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto,

Étiquette EMR de sortie Amazon	Version Presto	Composants installés avec Presto
		hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Le tableau suivant répertorie la version de Presto incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Presto.

Pour connaître la version des composants installés avec Presto dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version Presto pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Presto	Composants installés avec Presto
emr-6.15.0	Presto 0.283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Le tableau suivant répertorie la version de Presto incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Presto.

Pour la version des composants installés avec Presto dans cette version, voir la [version 5.36.2](#) Versions des composants.

Informations sur la version Presto pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Presto	Composants installés avec Presto
emr-5.36.2	Presto 0.267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Le tableau suivant répertorie la version de Trino (PrestoSQL) incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Trino (Presto). SQL

Pour la version des composants installés avec Trino (PrestoSQL) dans cette version, consultez la [version 6.15.0](#) Versions des composants.

Informations de version de Trino (PrestoSQL) pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.15.0	Trino (PrestoSQL) 426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-

Étiquette EMR de sortie Amazon	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
		server, mariadb-server, trino-coordinator, trino-worker

Rubriques

- [Utilisation de Presto avec le catalogue de données AWS Glue](#)
- [Utilisation de S3 Select Pushdown avec Presto pour améliorer les performances](#)
- [Ajout de connecteurs de base de données](#)
- [Utilisation de SSL/TLS configuration LDAPS avec Presto sur Amazon EMR](#)
- [Activation du mode strict Presto](#)
- [Gestion de la perte d'instances Spot dans Presto](#)
- [Exécution tolérante aux pannes dans Trino](#)
- [Utilisation du dimensionnement automatique de Presto avec désaffectation gracieuse](#)
- [Considérations relatives à Presto sur Amazon EMR](#)
- [Historique des versions de Presto](#)

Utilisation de Presto avec le catalogue de données AWS Glue

À l'aide des EMR versions 5.10.0 et ultérieures d'Amazon, vous pouvez spécifier le catalogue de données AWS Glue comme métastore Hive par défaut pour Presto. Cette configuration est recommandée si vous avez besoin d'un métastore persistant ou d'un métastore partagé par différents clusters, services, applications ou Comptes AWS.

AWS Glue est un service d'extraction, de transformation et de chargement (ETL) entièrement géré qui permet de classer vos données, de les nettoyer, de les enrichir et de les déplacer de manière fiable entre différents magasins de données de manière simple et rentable. Le catalogue de données AWS Glue fournit un référentiel de métadonnées unifié couvrant une variété de sources de données et de formats de données, s'intégrant à Amazon EMR ainsi qu'à AmazonRDS, Amazon Redshift, Redshift Spectrum, Athena et à toute application compatible avec le métastore Apache Hive. Les robots d'exploration Glue peuvent automatiquement déduire un schéma à partir des données sources dans Amazon S3 et stocker les métadonnées associées dans le catalogue de données. Pour

plus d'informations sur le catalogue de données, consultez la section [Remplissage du catalogue de données AWS Glue](#) dans le AWS Glue Developer Guide.

Des frais distincts s'appliquent pour AWS Glue. Il existe un tarif mensuel pour le stockage et l'accès aux métadonnées du catalogue de données, un taux horaire facturé par minute pour les ETL tâches AWS Glue et le temps d'exécution du robot d'exploration, et un taux horaire facturé par minute pour chaque point de terminaison de développement provisionné. Le catalogue de données vous permet de stocker jusqu'à un million d'objets sans aucuns frais. Si vous stockez plus d'un million d'objets, vous êtes facturé USD 1\$ pour 100 000 objets de plus d'un million. Dans le catalogue de données, sont considérés comme des objets une table, une partition ou une base de données. Pour de plus amples informations, consultez [Tarification Glue](#).

Important

Si vous avez créé des tables à l'aide d'Amazon Athena ou d'Amazon Redshift Spectrum avant le 14 août 2017, les bases de données et les tables sont stockées dans un catalogue géré par Athena, distinct du catalogue de données Glue. AWS Pour intégrer Amazon EMR à ces tables, vous devez passer au catalogue de données AWS Glue. Pour plus d'informations, consultez la section [Mise à niveau vers le catalogue de données AWS Glue](#) dans le guide de l'utilisateur Amazon Athena.

Spécifier le catalogue de données AWS Glue comme métastore

Vous pouvez spécifier le catalogue de données AWS Glue comme métastore à l'aide de AWS Management Console AWS CLI, ou Amazon. EMR API Lorsque vous utilisez le CLI ou API, vous utilisez la classification de configuration de Presto pour spécifier le catalogue de données. En outre, avec Amazon EMR 5.16.0 et versions ultérieures, vous pouvez utiliser la classification de configuration pour spécifier un catalogue de données dans un autre. Compte AWS Lorsque vous utilisez la console, vous pouvez spécifier le catalogue de données à l'aide des Options avancées ou des Options rapides.

Console

Pour spécifier AWS Glue Data Catalog comme métastore Hive avec la nouvelle console

1. Connectez-vous au et ouvrez AWS Management Console la EMR console Amazon à l'adresse <https://console.aws.amazon.com/emr>.

2. Sous EMREC2Activé dans le volet de navigation de gauche, choisissez Clusters, puis Create cluster.
3. Sous Solution groupée d'application, choisissez Presto.
4. Sous Paramètres du catalogue de données AWS Glue, cochez la case Utiliser pour les métadonnées des tables Presto.
5. Choisissez toutes les autres options qui s'appliquent à votre cluster.
6. Pour lancer cluster, choisissez Créer un cluster.

CLI

Pour spécifier le catalogue de données AWS Glue comme métastore Hive par défaut à l'aide du AWS CLI

Pour des exemples de spécification des classifications de configuration suivantes lorsque vous créez un cluster, consultez [Configuration des applications](#).

Amazon EMR 5.16.0 et versions ultérieures

- Définissez la `hive.metastore` propriété sur `glue` comme indiqué dans l'JSONexemple suivant.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

Pour spécifier un catalogue de données dans un autre Compte AWS, ajoutez la `hive.metastore.glue.catalogid` propriété comme indiqué dans l'JSONexemple suivant. Remplacez *acct-id* par le compte AWS du catalogue de données. L'utilisation d'un catalogue de données dans un autre n' Compte AWS est pas disponible dans les EMR versions 5.15.0 et antérieures d'Amazon.

```
[
  {
    "Classification": "presto-connector-hive",
```

```
"Properties": {
  "hive.metastore": "glue",
  "hive.metastore.glue.catalogid": "acct-id"
}
]
```

Amazon EMR 5.10.0 à 5.15.0

Définissez la `hive.metastore.glue.datacatalog.enabled` propriété sur `true`, comme indiqué dans l'JSONexemple suivant :

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore.glue.datacatalog.enabled": "true"
    }
  }
]
```

Amazon EMR 6.1.0 et versions ultérieures utilisant Presto SQL (Trino)

À partir de EMR la version 6.1.0, Presto prend SQL également en charge Glue en tant que métastore Hive par défaut. Utilisez la classification `prestosql-connector-hive` de configuration et définissez la `hive.metastore` propriété sur `glue`, comme indiqué dans l'JSONexemple suivant.

EMR Les versions 6.4.0 et ultérieures d'Amazon utilisent le nouveau nom Trino au lieu de Presto. SQL Si vous utilisez Trino, remplacez `prestosql-connector-hive` dans la classification de configuration suivante par `trino-connector-hive`.

```
[
  {
    "Classification": "prestosql-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

Pour basculer les metastores sur un cluster de longue durée, vous pouvez définir manuellement ces valeurs selon les besoins de votre version en vous connectant au nœud maître, en modifiant les valeurs de propriété directement dans le fichier `/etc/presto/conf/catalog/hive.properties` et en redémarrant le serveur Presto (`sudo restart presto-server`). Si vous utilisez cette méthode avec Amazon EMR 5.15.0 et versions antérieures, assurez-vous qu'elle `hive.table-statistics-enabled` est définie sur `false`. Ce paramètre n'est pas obligatoire lorsque vous utilisez les versions 5.16.0 et ultérieures ; cependant, les statistiques de table et de partition ne sont pas prises en charge.

IAM autorisations

Le profil d'EC2 instance d'un cluster doit disposer d'IAM autorisations pour les actions AWS Glue. En outre, si vous activez le chiffrement pour les objets AWS Glue Data Catalog, le rôle doit également être autorisé à chiffrer, déchiffrer et générer les objets AWS KMS key utilisés pour le chiffrement.

Autorisations pour les actions AWS Glue

Si vous utilisez le profil d'EC2 instance par défaut pour Amazon EMR, aucune action n'est requise. La politique `AmazonElasticMapReduceforEC2Role` gérée attachée au `EMR_EC2_DefaultRole` autorise toutes les actions AWS Glue nécessaires. Toutefois, si vous spécifiez un profil d'EC2 instance personnalisé et des autorisations, vous devez configurer les actions AWS Glue appropriées. Utilisez la stratégie gérée `AmazonElasticMapReduceforEC2Role` comme point de départ. Pour plus d'informations, consultez la section [Rôle de service pour les EC2 instances de cluster \(profil d'EC2 instance\)](#) dans le guide EMR de gestion Amazon.

Autorisations pour chiffrer et déchiffrer le catalogue de données Glue AWS

Votre profil d'instance a besoin d'autorisations pour chiffrer et déchiffrer les données au moyen de votre clé. Vous n'avez pas besoin de configurer ces autorisations si les deux déclarations suivantes s'appliquent :

- Vous activez le chiffrement des objets du catalogue de données AWS Glue à l'aide de clés gérées pour AWS Glue.
- Vous utilisez un cluster Compte AWS identique au AWS Glue Data Catalog.

Dans le cas contraire, vous devez ajouter l'instruction suivante à la politique d'autorisation attachée à votre profil d'EC2 instance.


```
[
  {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Action": [
          "kms:Decrypt",
          "kms:Encrypt",
          "kms:GenerateDataKey"
        ],
        "Resource": "arn:aws:kms:region:acct-id:key/12345678-1234-1234-1234-123456789012"
      }
    ]
  }
]
```

Pour plus d'informations sur le chiffrement du catalogue de données AWS Glue, consultez la section [Chiffrer votre catalogue de données](#) dans le guide du développeur de AWS Glue.

Autorisations basées sur les ressources

Si vous utilisez AWS Glue conjointement avec Hive, Spark ou Presto sur Amazon, EMR AWS Glue prend en charge les politiques basées sur les ressources afin de contrôler l'accès aux ressources du catalogue de données. Ces ressources comprennent les bases de données, les tables, les connexions et les fonctions définies par l'utilisateur. Pour plus d'informations, consultez [Politiques de ressources AWS Glue](#) dans le Guide du développeur AWS Glue.

Lorsque vous utilisez des politiques basées sur les ressources pour limiter l'accès à AWS Glue depuis AmazonEMR, le principal que vous spécifiez dans la politique d'autorisation doit être le rôle ARN associé au profil d'EC2instance spécifié lors de la création d'un cluster. Par exemple, pour une politique basée sur les ressources attachée à un catalogue, vous pouvez spécifier le rôle du rôle de service par défaut ARN pour les instances de clusterEC2, *EMR_EC2_DefaultRole* comme lePrincipal, en utilisant le format illustré dans l'exemple suivant :

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

Le *acct-id* peut être différent de l'identifiant du compte AWS Glue. Cela permet l'accès à partir de EMR clusters appartenant à différents comptes. Vous pouvez spécifier plusieurs principaux, chacun provenant d'un compte différent.

Considérations relatives à l'utilisation du catalogue de données AWS Glue

Tenez compte des éléments suivants lorsque vous utilisez AWS Glue Data Catalog comme métastore avec Presto :

- Il n'est pas possible de renommer des tables depuis AWS Glue.
- Lorsque vous créez une table Hive sans spécifier de LOCATION, les données de la table sont stockées à l'emplacement spécifié par la propriété `hive.metastore.warehouse.dir`. Par défaut, il s'agit d'un emplacement dans HDFS. Si un autre cluster doit accéder à la table, il échoue, sauf s'il dispose d'autorisations adéquates sur le cluster qui a créé la table. En outre, le HDFS stockage étant transitoire, si le cluster se termine, les données de la table sont perdues et la table doit être recréée. Nous vous recommandons de spécifier un LOCATION dans Amazon S3 lorsque vous créez une table Hive à l'aide de AWS Glue. Vous pouvez également utiliser la classification de configuration `hive-site` pour spécifier un emplacement dans Amazon S3 pour `hive.metastore.warehouse.dir`, qui s'applique à toutes les tables Hive. Si une table est créée dans un HDFS emplacement et que le cluster qui l'a créée est toujours en cours d'exécution, vous pouvez mettre à jour l'emplacement de la table vers Amazon S3 depuis AWS Glue. Pour plus d'informations, consultez [Working with Tables on the AWS Glue Console](#) dans le AWS Glue Developer Guide.
- Les valeurs de partition contenant des guillemets et des apostrophes ne sont pas prises en charge, par exemple `PARTITION (owner="Doe 's")`.
- Les [statistiques sur les colonnes](#) sont prises en charge à partir de la version emr-5.31.0.
- L'utilisation de l'[autorisation Hive](#) n'est pas prise en charge. Comme alternative, envisagez d'utiliser les [politiques basées sur les ressources AWS Glue](#). Pour plus d'informations, consultez [Utiliser des politiques basées sur les ressources pour le catalogue de données Amazon EMR Access to AWS Glue](#).

Utilisation de S3 Select Pushdown avec Presto pour améliorer les performances

Avec les EMR versions 5.18.0 et ultérieures d'Amazon, vous pouvez utiliser [S3 select](#) Pushdown avec Presto sur Amazon. Cette fonctionnalité permet à Presto de « pousser vers le bas » le

travail de calcul d'opérations de projection (par exemple, SELECT) et les opérations de prédicat (par exemple, WHERE) sur Amazon S3. Cela permet aux requêtes de récupérer uniquement les données requises d'Amazon S3, ce qui peut améliorer les performances et réduire la quantité de données transférées entre Amazon EMR et Amazon S3 dans certaines applications.

S3 Select Pushdown est-il adapté à mon application ?

Nous vous recommandons de tester vos applications avec et sans S3 Select Pushdown pour voir si son utilisation peut être adaptée à votre application.

Utilisez les consignes suivantes pour déterminer si votre application peut utiliser S3 Select :

- Votre requête exclut plus de la moitié de l'ensemble de données d'origine.
- Vos prédicats de filtre de requête utilisent les colonnes qui ont un type de données pris en charge par Presto et S3 Select. Les types de données horodateur, réel et double ne sont pas pris en charge par S3 Select Pushdown. Nous recommandons d'utiliser le type de données décimal pour les données quantitatives. Pour plus d'informations sur les types de données pris en charge pour S3 Select, veuillez consulter Amazon Simple Storage Service, veuillez consulter [Types de données](#) dans le Guide de l'utilisateur Amazon Simple Storage Service.
- Votre connexion réseau entre Amazon S3 et le EMR cluster Amazon offre une bonne vitesse de transfert et une bonne bande passante disponible. Amazon S3 ne compresse pas HTTP les réponses. La taille des réponses est donc susceptible d'augmenter pour les fichiers d'entrée compressés.

Considérations et restrictions

- Seuls les objets stockés au CSV format sont pris en charge. Les objets peuvent être décompressés ou facultativement compressés avec gzip ou bzip2.
- La propriété `AllowQuotedRecordDelimiters` n'est pas prise en charge. Si cette propriété est spécifiée, la requête échoue.
- Le chiffrement côté serveur Amazon S3 avec les clés de chiffrement fournies par le client (SSE-C) et le chiffrement côté client ne sont pas pris en charge.
- S3 Select Pushdown ne remplace pas l'utilisation de formats de fichiers en colonnes ou compressés tels que ParquetORC.

Activation de S3 Select Pushdown avec PrestoDB ou Trino

Pour activer S3 Select Pushdown pour PrestoDB sur EMR Amazon, utilisez la classification de configuration à `hive.s3select-pushdown.enabled` définir `presto-connector-hive true` comme indiqué dans l'exemple ci-dessous. Pour de plus amples informations, veuillez consulter [Configuration des applications](#). La valeur `hive.s3select-pushdown.max-connections` doit également être définie. Le paramètre `500` par défaut devrait être approprié pour la plupart des applications. Pour plus d'informations, consultez [Comprendre et régler `hive.s3select-pushdown.max-connections`](#) ci-dessous.

Pour Presto SQL sur EMR les versions 6.1.0 à 6.3.0, remplacez `presto-connector-hive` dans l'exemple ci-dessous par `presto-sql-connector-hive`

EMR Les versions 6.4.0 et ultérieures d'Amazon utilisent le nouveau nom Trino au lieu de Presto. SQL Si vous utilisez Trino, remplacez `presto-connector-hive` dans l'exemple ci-dessous par `trino-connector-hive`

```
[
  {
    "classification": "presto-connector-hive",
    "properties": {
      "hive.s3select-pushdown.enabled": "true",
      "hive.s3select-pushdown.max-connections": "500"
    }
  }
]
```

Comprendre et régler `hive.s3select-pushdown.max-connections`

Par défaut, Presto utilise EMRFS comme système de fichiers. Le paramètre `fs.s3.maxConnections` de la classification `emrfs-site` de configuration spécifie le nombre maximal de connexions client autorisées à Amazon S3 via EMRFS Presto. 500 par défaut. S3 Select Pushdown contourne les opérations de prédicat EMRFS lors de l'accès à Amazon S3. Dans ce cas, la valeur de `hive.s3select-pushdown.max-connections` détermine le nombre maximal de connexions clients autorisées pour ces opérations de nœuds de travail. Cependant, toutes les demandes adressées à Amazon S3 par Presto qui ne sont pas repoussées vers le bas (par exemple, les GET opérations) continuent d'être régies par la valeur de `fs.s3.maxConnections`

Si votre application fait l'expérience de l'erreur « Temporisation attente de connexion du groupe », augmentez la valeur de `hive.s3select-pushdown.max-connections` et de `fs.s3.maxConnections`.

Ajout de connecteurs de base de données

Vous pouvez utiliser les classifications de configuration pour configurer les propriétés du JDBC connecteur lorsque vous créez un cluster. Les classifications de configuration commencent par `presto-connector`, par exemple `presto-connector-postgresql`. Les classifications de configuration disponibles dépendent de la EMR version publiée par Amazon. Pour les classifications de configuration disponibles avec la version la plus récente, consultez [the section called "5.36.2 classifications de configuration"](#) Amazon EMR 5.36.2. Si vous utilisez une autre version d'AmazonEMR, consultez [Versions publiées par Amazon EMR 5.x](#) les classifications de configuration. Pour plus d'informations sur les propriétés qui peuvent être configurées avec chaque connecteur, consultez <https://prestodb.io/docs/current/connector.html>.

Exemple —configuration d'un cluster avec le connecteur Postgre SQL JDBC

Pour lancer un cluster avec le SQL connecteur Postgre installé et configuré, créez d'abord un JSON fichier qui spécifie la classification de configuration, par exemple, `myConfig.json` avec le contenu suivant, et enregistrez-le localement.

Remplacez les propriétés de connexion en fonction de votre configuration et comme indiqué dans la rubrique relative au [SQLconnecteur Postgre](#) de la documentation Presto.

```
[
  {
    "Classification": "presto-connector-postgresql",
    "Properties": {
      "connection-url": "jdbc:postgresql://example.net:5432/database",
      "connection-user": "MYUSER",
      "connection-password": "MYPASS"
    },
    "Configurations": []
  }
]
```

Lorsque vous créez le cluster, référencez le chemin du JSON fichier à l'aide de l'--`configurationsoption` illustrée dans l'exemple suivant, où se `myConfig.json` trouve le répertoire dans lequel vous exécutez la commande :

```
aws emr create-cluster --name PrestoConnector --release-label emr-5.36.2 --instance-type m5.xlarge \  
--instance-count 2 --applications Name=Hadoop Name=Hive Name=Pig Name=Presto \  
--use-default-roles --ec2-attributes KeyName=myKey \  
--log-uri s3://my-bucket/logs --enable-debugging \  
--configurations file://myConfig.json
```

Utilisation de SSL/TLS configuration LDAPS avec Presto sur Amazon EMR

Avec les EMR versions 5.6.0 et ultérieures d'Amazon, vous pouvez SSL activer/ TLS pour [sécuriser les communications internes](#) entre les nœuds Presto. Pour ce faire, vous devez définir une configuration de sécurité pour le chiffrement en transit. Pour plus d'informations, consultez les [sections Options de chiffrement](#) et [Utiliser les configurations de sécurité pour configurer la sécurité du cluster](#) dans le Amazon EMR Management Guide.

Lorsque vous utilisez une configuration de sécurité avec chiffrement en transit, Amazon EMR effectue les opérations suivantes pour Presto :

- Il distribue les artefacts de chiffrement, ou certificats, que vous spécifiez pour le chiffrement en transit sur l'ensemble du cluster Presto. Pour plus d'informations, consultez [Mise à disposition des certificats de chiffrement des données en transit](#).
- Il définit les propriétés suivantes à l'aide de la classification de configuration presto-config, qui correspond au fichier config.properties pour Presto :
 - Définit http-server.http.enabled la valeur false sur tous les nœuds, ce qui la désactive HTTP en faveur de HTTPS. Cela nécessite que vous fournissiez des certificats qui fonctionnent pour le public et le privé DNS lors de la configuration de sécurité pour le chiffrement en transit. Pour ce faire, vous pouvez utiliser des certificats SAN (nom alternatif du sujet) qui prennent en charge plusieurs domaines.
 - Il définit les valeurs http-server.https.*. Pour plus de détails sur la configuration, consultez [LDAP/ authentification](#) dans la documentation Presto.
- Pour Presto SQL (Trino) sur les EMR versions 6.1.0 et ultérieures, Amazon configure EMR automatiquement une clé secrète partagée pour une communication interne sécurisée entre les nœuds du cluster. Vous n'avez pas besoin d'effectuer de configuration supplémentaire pour activer cette fonctionnalité de sécurité, et vous pouvez remplacer la configuration par votre propre clé

secrète. Pour plus d'informations sur l'authentification interne de Trino, consultez la [documentation Trino 353 : Communication interne sécurisée](#).

En outre, avec les EMR versions 5.10.0 et ultérieures d'Amazon, vous pouvez configurer l'[LDAPauthentification](#) pour les connexions des clients au coordinateur Presto à l'aide de HTTPS. Cette configuration utilise secure LDAP (LDAPS). TLS doit être activé sur votre LDAP serveur, et le cluster Presto doit utiliser une configuration de sécurité avec le chiffrement des données en transit activé. Une configuration supplémentaire est requise. Les options de configuration sont différentes selon la version d'Amazon EMR que vous utilisez. Pour de plus amples informations, veuillez consulter [Utilisation de LDAP l'authentification pour Presto sur Amazon EMR](#).

Presto sur Amazon EMR utilise le port 8446 pour le mode interne HTTPS par défaut. Le port utilisé pour la communication interne doit être le même que celui utilisé pour l'HTTPS accès du client au coordinateur Presto. La propriété `http-server.https.port` dans la classification de configuration `presto-config` spécifie le port.

Utilisation de LDAP l'authentification pour Presto sur Amazon EMR

Suivez les étapes décrites dans cette section pour configurer LDAP. Chaque étape contient des exemples et des liens vers des informations complémentaires.

Étapes de configuration de LDAP l'authentification

- [Étape 1 : recueillir des informations sur votre LDAP serveur et copier le certificat du serveur sur Amazon S3](#)
- [Étape 2 : Définir une configuration de sécurité](#)
- [Étape 3 : Création d'une configuration JSON avec les propriétés Presto pour LDAP](#)
- [Étape 4 : créer le script pour copier le certificat LDAP du serveur et le télécharger sur Amazon S3](#)
- [Étape 5 : Créer le cluster](#)

Étape 1 : recueillir des informations sur votre LDAP serveur et copier le certificat du serveur sur Amazon S3

Vous aurez besoin des informations et des éléments de la section suivante auprès de votre LDAP serveur pour configurer LDAP l'authentification.

Adresse IP ou nom d'hôte du LDAP serveur

Le coordinateur Presto sur le nœud EMR principal Amazon doit pouvoir accéder au LDAP serveur à l'adresse IP ou au nom d'hôte spécifiés. Par défaut, Presto communique avec le LDAP serveur LDAPS via le port 636. Si votre LDAP implémentation nécessite un port personnalisé, vous pouvez le spécifier à l'aide de la `ldap.url` propriété avec Amazon EMR 5.16.0 ou version ultérieure, ou en utilisant `authentication.ldap.url` des versions antérieures. Remplacez 636 par le port personnalisé comme illustré dans les exemples de configuration `presto-config` de [Étape 3 : Création d'une configuration JSON avec les propriétés Presto pour LDAP](#). Assurez-vous que les pare-feu et les groupes de sécurité autorisent le trafic entrant et sortant sur le port 636 (ou votre port personnalisé) ainsi que sur le port 8446 (ou votre port personnalisé), qui est utilisé pour les communications de cluster internes.

Le certificat LDAP du serveur

Vous devez charger le fichier de certificat dans un emplacement sécurisé dans Amazon S3. Pour plus d'informations, consultez [Comment charger des fichiers ou des dossiers dans un compartiment S3](#) dans le Guide de l'utilisateur Amazon Simple Storage Service. Vous créez une action d'amorçage qui copie ce certificat depuis Amazon S3 vers chaque nœud du cluster lors du lancement du cluster. Dans [Étape 4 : créer le script pour copier le certificat LDAP du serveur et le télécharger sur Amazon S3](#). L'exemple de certificat est `s3://MyBucket/ldap_server.crt`.

Les paramètres du LDAP serveur pour la liaison anonyme

Si la liaison anonyme est désactivée sur PrestoDB, vous avez besoin de l'ID utilisateur UID () et du mot de passe d'un compte autorisé à établir une liaison au serveur afin que le LDAP serveur PrestoDB puisse établir une connexion. Vous spécifiez le mot de passe UID et à l'aide `internal-communication.authentication.ldap.password` des propriétés `internal-communication.authentication.ldap.user` et de la classification `presto-config` de configuration. Amazon EMR 5.10.0 ne prend pas en charge ces paramètres. La liaison anonyme doit donc être prise en charge sur le LDAP serveur lorsque vous utilisez cette version.

Notez que Trino ne nécessite pas la configuration de liaison anonyme.

Pour obtenir l'état de la liaison anonyme sur le LDAP serveur

- Utilisez la commande [ldapwhoami](#) à partir d'un client Linux, comme illustré dans l'exemple suivant :

```
ldapwhoami -x -H ldaps://LDAPServerHostNameOrIPAddress
```


Si la liaison anonyme n'est pas autorisée, la commande renvoie les informations suivantes :

```
ldap_bind: Inappropriate authentication (48)
additional info: anonymous bind disallowed
```

Pour vérifier qu'un compte est autorisé à accéder à un LDAP serveur qui utilise une authentification simple

- Utilisez la commande [ldapwhoami](#) à partir d'un client Linux, comme illustré dans l'exemple suivant. L'exemple utilise un utilisateur fictif, *presto*, stocké sur un LDAP serveur Open exécuté sur une EC2 instance portant le nom d'hôte fictif *ip-xxx-xxx-xxx-xxx.ec2.internal*. L'utilisateur est associé à l'unité organisationnelle (UO) *admins* et avec le mot de passe *123456*:

```
ldapwhoami -x -w "123456" -D uid=presto,ou=admins,dc=ec2,dc=internal -H ldaps://ip-
xxx-xxx-xxx-xxx.ec2.internal
```

Si le compte est valide et dispose des autorisations appropriées, la commande renvoie :

```
dn:uid=presto,ou=admins,dc=ec2,dc=internal
```

Les exemples de configuration dans [Étape 3 : Création d'une configuration JSON avec les propriétés Presto pour LDAP](#) incluent ce compte pour plus de clarté, à l'exception de l'exemple 5.10.0, dans lequel il n'est pas pris en charge. Si le LDAP serveur utilise une liaison anonyme, supprimez les paires `internal-communication.authentication.ldap.password nom/valeur` `internal-communication.authentication.ldap.user` et.

Le nom LDAP distinctif (DN) pour les utilisateurs de Presto

Lorsque vous spécifiez la LDAP configuration de Presto, vous spécifiez un modèle de liaison composé de `${USER}` une unité organisationnelle (UO) et de composants de domaine supplémentaires (DCs). Presto remplace `${USER}` par l'ID utilisateur (UID) réel de chaque utilisateur lors de l'authentification par mot de passe afin qu'il corresponde au nom distinctif (DN) spécifié par ce modèle de liaison. Vous avez besoin de OUs ceux auxquels appartiennent les utilisateurs éligibles et de leurs DCs. Par exemple, pour autoriser les utilisateurs de l'unité d'organisation

admins ou du domaine corp.example.com à s'authentifier auprès de Presto, vous spécifiez `${USER},ou=admins,dc=corp,dc=example,dc=com` en tant que modèle de liaison utilisateur.

Note

Lorsque vous l'utilisez AWS CloudFormation, vous devez utiliser la fonction Fn::Sub afin de la `${USER}` remplacer par l'ID utilisateur réel (UID). Pour plus d'informations, consultez la rubrique [Fn::Sub](#) dans le Guide de l'utilisateur AWS CloudFormation .

Lorsque vous utilisez Amazon EMR 5.10.0, vous ne pouvez spécifier qu'un seul modèle de ce type. À l'aide d'Amazon EMR 5.11.0 ou version ultérieure, vous pouvez spécifier plusieurs modèles séparés par deux points (:). Les utilisateurs qui tentent de s'authentifier auprès de Presto sont comparés au premier modèle, puis au deuxième, et ainsi de suite. Pour obtenir un exemple, consultez [Étape 3 : Création d'une configuration JSON avec les propriétés Presto pour LDAP](#).

Étape 2 : Définir une configuration de sécurité

Créez une configuration de sécurité avec le chiffrement en transit activé. Pour plus d'informations, consultez la section [Création d'une configuration de sécurité](#) dans le guide EMR de gestion Amazon. Les artefacts de chiffrement que vous fournissez lorsque vous configurez le chiffrement en transit sont utilisés pour chiffrer les communications internes entre les nœuds Presto. Pour plus d'informations, consultez [Mise à disposition des certificats de chiffrement des données en transit](#). Le certificat de LDAP serveur est utilisé pour authentifier les connexions des clients au serveur Presto.

Étape 3 : Création d'une configuration JSON avec les propriétés Presto pour LDAP

Vous utilisez la classification `presto-config` de configuration pour définir les propriétés Presto pourLDAP. Le format et le contenu de `presto-config` sont légèrement différents selon la EMR version d'Amazon et l'installation de Presto (PrestoDB ou Trino). Des exemples de configuration sont fournis plus tard dans cette section. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

Les étapes suivantes supposent que vous enregistrez les JSON données dans un fichier, *MyPrestoConfig.json*. Si vous utilisez la console, chargez le fichier sur un emplacement sécurisé dans Amazon S3 afin de pouvoir le référencer lors de la création du cluster. Si vous utilisez le AWS CLI, vous pouvez référencer le fichier localement.

Exemple Amazon EMR 6.1.0 et versions ultérieures avec Presto SQL (Trino)

L'exemple suivant utilise le nom LDAP d'hôte de [Étape 1 : recueillir des informations sur votre LDAP serveur et copier le certificat du serveur sur Amazon S3](#) pour s'authentifier auprès LDAP du serveur à des fins de liaison. Deux modèles de liaison utilisateur sont spécifiés, ce qui indique que les utilisateurs au sein de l'unité admins d'organisation et de l'unitéd'organisation sur le LDAP serveur sont éligibles à l'authentification auprès du serveur Trino en tant qu'utilisateurs. Les modèles de liaison sont séparés par un signe deux-points (:).

EMR Les versions 6.4.0 et ultérieures d'Amazon utilisent le nouveau nom Trino au lieu de Presto. SQL Si vous utilisez Trino, remplacez *configuration de prestosql-config* dans la classification de configuration suivante avec *trino-config* et *prestosql-password-authenticator* avec *trino-password-authenticator*.

```
[
  {
    "Classification": "prestosql-config",
    "Properties": {
      "http-server.authentication.type": "PASSWORD"
    }
  },
  {
    "Classification": "prestosql-password-authenticator",
    "Properties": {
      "password-authenticator.name": "ldap",
      "ldap.url": "ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",
      "ldap.user-bind-pattern": "uid=${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal"
    }
  }
]
```

Exemple Amazon EMR 5.16.0 et versions ultérieures

L'exemple suivant utilise l'ID LDAP utilisateur et le mot de passe, ainsi que le nom LDAP d'hôte de [Étape 1 : recueillir des informations sur votre LDAP serveur et copier le certificat du serveur sur Amazon S3](#) pour s'authentifier auprès LDAP du serveur à des fins de liaison. Deux modèles de liaison utilisateur sont spécifiés, ce qui indique que les utilisateurs au sein de l'unité admins d'organisation et de l'unitéd'organisation sur le LDAP serveur sont éligibles à l'authentification auprès du serveur Presto en tant qu'utilisateurs. Les modèles de liaison sont séparés par un signe deux-points (:).

```
[{
  "Classification": "presto-config",
  "Properties": {
    "http-server.authentication.type": "PASSWORD"
  }
},
{
  "Classification": "presto-password-authenticator",
  "Properties": {
    "password-authenticator.name": "ldap",
    "ldap.url": "ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal:636",
    "ldap.user-bind-pattern": "uid=
${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal",
    "internal-communication.authentication.ldap.user": "presto",
    "internal-communication.authentication.ldap.password": "123456"
  }
}]
```

Exemple Amazon EMR 5.11.0 à 5.15.0

Le format de la classification de configuration `presto-config` est légèrement différent pour ces versions. L'exemple suivant spécifie les mêmes paramètres que l'exemple précédent.

```
[{
  "Classification": "presto-config",
  "Properties": {
    "http-server.authentication.type": "LDAP",
    "authentication.ldap.url": "ldaps://ip-xxx-xxx-xxx-
xxx.ec2.internal:636",
    "authentication.ldap.user-bind-pattern": "uid=
${USER},ou=admins,dc=ec2,dc=internal:uid=${USER},ou=datascientists,dc=ec2,dc=internal",
    "internal-communication.authentication.ldap.user": "presto",
    "internal-communication.authentication.ldap.password": "123456"
  }
}]
```

Exemple Amazon EMR 5.10.0

Amazon EMR 5.10.0 prend uniquement en charge la liaison anonyme, ces entrées sont donc omises. De plus, un seul modèle de liaison peut être spécifié.

```
[{
```

```

    "Classification": "presto-config",
      "Properties": {
        "http-server.authentication.type": "LDAP",
        "authentication.ldap.url": "ldaps://ip-xxx-xxx-xxx-
xxx.ec2.internal:636",
        "ldap.user-bind-pattern": "uid=
${USER},ou=prestousers,dc=ec2,dc=internal"
      }
    ]
  ]
}

```

Étape 4 : créer le script pour copier le certificat LDAP du serveur et le télécharger sur Amazon S3

Créez un script qui copie le fichier de certificat sur chaque nœud du cluster et l'ajoute au keystore. Créez le script à l'aide d'un éditeur de texte, enregistrez-le, puis chargez-le dans Amazon S3.

Dans [Étape 5 : Créer le cluster](#), le fichier de script est référencé comme `s3://MyBucket/LoadLDAPCert.sh`.

L'exemple de script suivant utilise le mot de passe du keystore par défaut, `changeit`. Nous vous recommandons de vous connecter au nœud principal après avoir créé le cluster et modifié le mot de passe du keystore à l'aide de la commande `keytool`.

```

#!/bin/bash
aws s3 cp s3://MyBucket/ldap_server.crt .
sudo keytool -import -keystore /usr/lib/jvm/jre-1.8.0-openjdk.x86_64/lib/security/
cacerts -trustcacerts -alias ldap_server -file ./ldap_server.crt -storepass changeit -
noprompt

```

Étape 5 : Créer le cluster

Lorsque vous créez le cluster, vous spécifiez Presto et les autres applications que vous souhaitez qu'Amazon EMR installe. Les exemples suivants font également référence aux propriétés de classification de configuration au sein d'un JSON, mais vous pouvez également spécifier la classification de configuration en ligne.

Pour créer un cluster Presto avec LDAP authentication à l'aide de la console Amazon EMR

1. Accédez à la nouvelle EMR console Amazon et sélectionnez [Basculer vers l'ancienne console](#) dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).

2. Choisissez Créer un cluster et Go to advanced options (Aller aux options avancées).
3. Choisissez Presto ainsi que d'autres applications EMR à installer par Amazon, puis sous Configuration logicielle, sélectionnez la version d'Amazon EMR à utiliser. LDAP l'authentification n'est prise en charge qu'avec Amazon EMR 5.10.0 et versions ultérieures.
4. Sous Modifier les paramètres du logiciel, choisissez Charger JSON depuis S3, entrez l'emplacement dans Amazon S3 du fichier de JSON configuration que vous avez créé [Étape 3 : Création d'une configuration JSON avec les propriétés Presto pour LDAP](#), puis choisissez Next.
5. Configurez le matériel et la mise en réseau du cluster, puis choisissez Next (Suivant).
6. Choisissez Bootstrap Actions (Actions d'amorçage). Pour Add bootstrap action (Ajouter une action d'amorçage), sélectionnez Custom action (Action personnalisée), puis choisissez Configure and add (Configurer et ajouter).
7. Entrez un nom pour l'action bootstrap, entrez l'emplacement du script dans lequel vous l'avez créé [Étape 4 : créer le script pour copier le certificat LDAP du serveur et le télécharger sur Amazon S3](#), par exemple s3://MyBucket/LoadLDAPCert.sh, puis choisissez Ajouter.
8. Sous General Options (Options générales), Tags (Balises) et Additional Options (Options supplémentaires), choisissez les paramètres appropriés à votre application, puis choisissez Next (Suivant).
9. Choisissez Authentication and encryption (Authentification et chiffrement), puis sélectionnez la configuration de sécurité que vous avez créée dans [Étape 2 : Définir une configuration de sécurité](#).
10. Choisissez d'autres options de sécurité en fonction de votre application, puis choisissez Create cluster (Créer le cluster).

Pour créer un cluster Presto avec LDAP authentification à l'aide du AWS CLI

- Utilisez la commande `aws emr create-cluster`. Au minimum, spécifiez l'application Presto, ainsi que la classification de configuration Presto, le script d'amorçage et la configuration de sécurité que vous avez créée au cours des étapes précédentes. L'exemple suivant fait référence au fichier de configuration en tant que JSON fichier enregistré dans le répertoire où vous exécutez la commande. Le script d'amorçage, d'autre part, doit être enregistré dans Amazon S3. L'exemple suivant utilise `s3://MyBucket/LoadLDAPCert.sh`.

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

```
aws emr create-cluster --applications Name=presto --release-label emr-5.16.0 \  
--use-default-roles --ec2-attributes KeyName=MyKeyPair,SubnetId=subnet-1234ab5 \  
\ --instance-count 3 --instance-type m5.xlarge --region us-west-2 --name \  
"MyPrestoWithLDAPAuth" \  
--bootstrap-actions Name="Distribute LDAP server cert",Path="s3://MyBucket/  
LoadLDAPCert.sh" \  
--security-configuration MyPrestoLDAPSecCfg --configurations file://  
MyPrestoConfig.json
```

Activation du mode strict Presto

Dans certaines situations, les requêtes de longue durée peuvent entraîner des coûts élevés et inciter Amazon EMR à utiliser davantage de ressources de cluster. Cela réduit les ressources des autres charges de travail du cluster. Avec EMR les versions 6.8 et ultérieures d'Amazon, vous pouvez utiliser une fonctionnalité de mode strict qui rejette ou vous avertit des types de requêtes de longue durée suivants :

- Requêtes sans prédicats sur les colonnes partitionnées qui entraînent l'analyse de grandes quantités de données dans les tables
- Requêtes avec jointures croisées entre deux grandes tables
- Requêtes qui trient un grand nombre de lignes sans limite

Une fois que Presto a complètement optimisé le plan de requêtes, le mode strict s'exécute. Pour utiliser et personnaliser le mode strict en fonction des besoins de vos requêtes, vous pouvez configurer Presto de la manière suivante.

Configurations Presto pour le mode strict

Paramètre	Description	Par défaut
<code>strict-mode-enabled</code>	Active et désactive le mode strict. La valeur de <code>true</code> indique que le mode strict est activé.	<code>false</code>
<code>strict-mode-fail-query</code>	Rejette les requêtes si le mode strict détecte des requêtes probables de longue durée. Si <code>false</code> , Amazon EMR émet uniquement un avertissement.	<code>false</code>
<code>strict-mode-restrictions</code>	Spécifie les restrictions à appliquer lorsque le mode strict est activé. Le mode strict prend en charge les restrictions suivantes : <code>MANDATORY_PARTITION _ DISALLOW _ PREDICATEJOIN</code> , <code>CROSS _ _</code> et <code>LIMITED _ SORT</code> .	<code>MANDATORY_PARTITION_PREDICATE,DISALLOW_CROSS_JOIN,LIMITED_SORT</code>

Pour tester le mode strict, vous pouvez remplacer ces configurations ou les définir comme propriétés de session lorsque vous utilisez le client Presto.

Pour définir la configuration lors de la création du cluster à l'aide du AWS Management Console

1. Choisissez `Create cluster` et sélectionnez Amazon EMR version 6.8.0, puis Presto ou Trino. Pour plus d'informations, veuillez consulter [Installation de PrestoDB et Trino](#).
2. Spécifiez directement les propriétés de configuration pour le mode strict ou chargez un JSON fichier sur Amazon S3. Vous pouvez éventuellement sélectionner le pour votre métastore. Spécifiez vos sous-réseaux VPC, vos actions d'amorçage, votre paire de clés et votre groupe de sécurité. Choisissez `Créer un cluster` pour créer votre cluster.
3. Connectez-vous au nœud primaire du cluster et exécutez `presto-cli` ou `trino-cli`.

4. Soumettez vos questions. Le mode strict valide chaque requête et détermine si elle est de longue durée. En fonction de vos `strict-mode-fail-query` paramètres, Amazon EMR rejette la requête ou émet un avertissement.
5. Lorsque vous avez terminé vos requêtes, résiliez le cluster et supprimez vos ressources.

Pour définir la configuration sur un cluster en cours d'exécution à l'aide du AWS CLI

1. Connectez-vous au nœud principal de votre cluster avec le AWS CLI et run `presto-cli` `outrino-cli`.
2. Exécutez les commandes suivantes avec les valeurs souhaitées.

```
set session strict_mode_enabled = true;
set session strict_mode_fail_query = false;
set session strict_mode_restrictions = 'DISALLOW_CROSS_JOIN,LIMITED_SORT';
```

Considérations

Lorsque vous utilisez le mode strict, tenez compte des éléments suivants :

- Dans certains cas, le mode strict peut rejeter les requêtes de courte durée qui ne consomment pas beaucoup de ressources. Par exemple, les requêtes sur de petites tables n'appliquent pas de filtrage dynamique et ne remplacent pas les jointures internes par des jointures croisées. Cela peut amener la requête à utiliser le prédicat de partition obligatoire ou à interdire la jointure croisée. Dans ce cas, le mode strict rejette la requête.
- Le contrôle du mode strict n'est appliqué qu'aux types de EXPLAIN ANALYZE requête SELECT INSERTSELECT, CREATE TABLE AS et.
- Vous ne pouvez utiliser la restriction sur le prédicat de partition obligatoire qu'avec le connecteur Hive.

Gestion de la perte d'instances Spot dans Presto

Avec les instances Spot d'AmazonEMR, vous pouvez exécuter des charges de travail de mégadonnées sur la EC2 capacité inutilisée d'Amazon à moindre coût. En échange d'une réduction des coûts, Amazon EC2 peut interrompre les instances Spot avec une notification de deux minutes. Lorsque vous résiliez un nœud, Presto peut prendre jusqu'à 10 minutes avant qu'il ne renvoie une

erreur. Cela entraîne des retards inutiles dans les rapports d'erreur et les éventuelles nouvelles tentatives. La résiliation rapide est une fonctionnalité qui vous permet de contrôler la façon dont Presto gère les nœuds résiliés.

Le rôle du coordinateur Presto est de suivre tous les composants master en interrogeant régulièrement leur statut. Sans interruption rapide, le coordinateur ne consulte pas YARN NodeManager le statut de chaque nœud. Cela peut entraîner une longue boucle de nouvelles tentatives avant l'échec de la requête. En cas de résiliation rapide, le coordinateur Presto consulte l'état du nœud NodeManager dès que le sondage ne parvient pas à atteindre l'hôte. S' NodeManager il s'avère que le nœud est inactif, Presto abandonne toute nouvelle tentative, échoue à la requête et renvoie une erreur. `NODE_DECOMMISSIONED`

L'ensemble de paramètres de configuration suivant vous permet de contrôler et de personnaliser le comportement de Presto en cas de résiliation du nœud.

Configurations Presto pour la gestion des défaillances des nœuds

Paramètre	Description	Par défaut
<code>query.remote-task.max-backoff-duration</code>	Durée pendant laquelle le coordinateur continue de tenter de récupérer l'état de la tâche distante à partir des composants master.	10 minutes
<code>query.remote-task.quick-terminate-on-de-failure</code>	Active la défaillance rapide du nœud si le coordinateur ne peut pas atteindre le nœud ou ne peut pas se connecter au travailleur qui s'exécute sur ce nœud. La valeur de <code>query.remote-task.terminate-on-connect-exception</code> détermine si le coordinateur doit atteindre le nœud ou se connecter au travailleur.	true

Paramètre	Description	Par défaut
	<p>Le nœud échoue à la requête et Amazon le EMR supprime de la liste des travailleurs disponibles. Dans ce cas, vous ne pouvez pas utiliser le nœud pour planifier de nouvelles requêtes.</p> <p>Lorsque vous définissez cette valeur sur <code>false</code>, Presto revient à son comportement précédent où le coordinateur Presto essaie à nouveau d'atteindre le nœud (pour <code>query.remote-task.max-backoff-duration</code>) avant de marquer le nœud comme indisponible et d'échouer à la requête en cours sur le nœud.</p>	
<code>query.remote-task.terminate-on-connection-exception</code>	<p>Spécifie si Amazon EMR doit créer un nœud si l'hôte est accessible mais que le coordinateur ne parvient pas à se connecter au processus de travail de l'hôte. Lorsque vous définissez cette valeur sur <code>true</code>, vous activez l'échec rapide de la requête si l'hôte est inaccessible.</p>	<p><code>false</code></p>

Exécution tolérante aux pannes dans Trino

L'exécution tolérante aux pannes est un mécanisme de Trino qu'un cluster peut utiliser pour atténuer les échecs de requêtes. Pour ce faire, il réessaie les requêtes ou les tâches qui les composent lorsqu'elles échouent. Lorsque l'exécution tolérante aux pannes est activée, les données d'échange intermédiaires sont mises en attente et un autre utilisateur peut les réutiliser en cas de panne ou de toute autre erreur lors de l'exécution de la requête.

Pour plus d'informations sur l'exécution tolérante aux pannes dans Trino, voir [Project Tardigrade fournit ETL à la vitesse de Trino aux premiers utilisateurs](#) sur le blog de Trino.

Configuration

L'exécution tolérante aux pannes est désactivée par défaut. Pour activer la fonctionnalité, définissez la propriété de configuration `retry-policy` dans la classification `trino-config` sur QUERY ou TASK en fonction de la politique de réessai souhaitée, comme suit.

```
{"classification":
  "trino-config",
"properties":
  {
    "retry-policy":
      "QUERY"
  }
}
```

Une politique de nouvelle tentative **QUERY** indique à Trino de réessayer automatiquement une requête lorsqu'une erreur se produit sur un composant master. Nous vous recommandons d'utiliser une politique de QUERY nouvelles tentatives lorsque la majeure partie de la charge de travail du cluster Trino comprend de nombreuses petites requêtes.

Une politique de nouvelle tentative **TASK** indique à Trino de réessayer des tâches de requête individuelles en cas d'échec. Nous recommandons cette politique lorsque Trino exécute des requêtes par lots de grande taille. Le cluster peut réessayer de manière plus efficace des tâches plus petites au sein de la requête plutôt que de réessayer l'ensemble de la requête.

Responsable des échanges

Un gestionnaire d'échange stocke et gère les données mises en bobine pour une exécution tolérante aux pannes. Il utilise un stockage externe pour stocker les données déversées au-delà de la taille

de la mémoire tampon. Vous pouvez configurer un gestionnaire d'échange basé sur un système de fichiers qui stocke les données spoolées dans un emplacement spécifié, tel qu'Amazon S3, les systèmes compatibles avec Amazon S3 ou. HDFS

Les EMR versions 6.9.0 et ultérieures d'Amazon incluent la `trino-exchange-manager` classification permettant de configurer le gestionnaire d'échange. Ces versions prennent également en charge HDFS le spooling.

Configuration du gestionnaire d'échange

Utilisez la classification de configuration `trino-exchange-manager` pour configurer un gestionnaire d'échange. Cette classification crée en interne un fichier de configuration `etc/exchange-manager.properties` sur le coordinateur et tous les nœuds de travail. La classification définit également la propriété de configuration `exchange-manager.name` sur `filesystem`.

Par défaut, les EMR versions 6.9.0 et ultérieures d'Amazon sont utilisées HDFS comme gestionnaire d'échange. HDFS est disponible dans les EMR EC2 clusters Amazon, et le spooling se produit par défaut dans le `trino-exchange/` répertoire. Pour utiliser les paramètres par défaut, définissez la configuration suivante :

```
{"Classification":  
  "trino-exchange-manager"  
}
```

Si vous souhaitez fournir un emplacement personnalisé, définissez les propriétés suivantes dans la classification `trino-exchange-manager` :

- Définissez `exchange.use-local-hdfs` sur `true`.
- Définissez `exchange.base-directories` l'emplacement du répertoire personnalisé dans HDFS, par exemple, `exchange.base-directories=/exchange`. Si le répertoire personnalisé n'y figure pas déjà HDFS, Amazon le EMR créera.

HDFS configurations du gestionnaire d'échange

Sur la base des résultats de tests internes, nous vous recommandons de passer au mode local HDFS pour obtenir de meilleures performances de requête par rapport aux autres systèmes de fichiers basés sur le cloud. Vous pouvez définir les configurations suivantes pour le gestionnaire d'échange avec HDFS.

Configuration	Description	Paramètre par défaut
<code>exchange.hdfs.block-size</code>	Taille des blocs pour le HDFS stockage	4 Mo
<code>hdfs.config.resources</code>	Liste des chemins de fichiers à configurer HDFS	Dans <code>exchange.use-local-hdfs</code> est <code>true</code> , utilise les chemins d'accès à <code>core-site.xml</code> , aux fichiers <code>hdfs-site.xml</code> ; dans le cas contraire <code>null</code>

Pour d'autres propriétés de configuration d'exécution tolérantes aux pannes et pour savoir comment configurer Amazon S3 ou d'autres systèmes compatibles avec Amazon S3 pour le spoupage, consultez la page d'[exécution tolérante aux pannes](#) de la documentation de Trino.

Considérations et restrictions

- Si vous activez l'exécution tolérante aux pannes, cela désactive les opérations `write` pour les connecteurs qui ne prennent pas en charge `write` lorsque `retry-policy` est défini. Depuis la EMR version 6.9.0 d'Amazon, les connecteurs Delta Lake, Hive et Iceberg prennent en charge `write` les opérations avec `retry-policy`
- Si vous utilisez Exchange Manager et que vous effectuez des opérations d'E/S coûteuses, les performances de vos requêtes risquent de se dégrader tandis qu'Exchange Manager envoie les données intermédiaires vers un stockage externe.

Utilisation du dimensionnement automatique de Presto avec désaffectation gracieuse

Les EMR versions 5.30.0 et ultérieures d'Amazon incluent une fonctionnalité que vous pouvez utiliser pour définir une période de grâce pour certaines actions de dimensionnement. La période de grâce permet aux tâches Presto de continuer à s'exécuter avant que le nœud ne se résilie en raison d'une action de redimensionnement ou d'une demande de politique de dimensionnement automatique. Pour plus d'informations sur les règles de dimensionnement, consultez [Comprendre les règles de dimensionnement automatique](#) dans le Amazon EMR Management Guide. Presto autoscaling avec la

mise hors service gracieuse empêche la planification de nouvelles tâches sur un nœud en cours de mise hors service, tout en permettant aux tâches déjà en cours d'exécution de se terminer avant que le délai d'arrêt ne soit atteint. Les requêtes en cours d'exécution se termineront avant que le nœud soit désaffecté. Autoscaling n'est pas pris en charge sur les parcs d'instances.

Vous pouvez contrôler le délai d'exécution des tâches Presto après réception d'une demande d'arrêt automatique. Par défaut, le délai d'arrêt d'Amazon EMR est de 0 quelques minutes, ce qui signifie qu'Amazon met EMR immédiatement fin au nœud et à toutes les tâches Presto exécutées sur celui-ci, si une demande d'extension l'exige. Pour définir un délai d'expiration plus long pour les tâches Presto sur Amazon EMR afin de permettre aux requêtes en cours d'exécution de se terminer avant de réduire la taille d'un cluster, utilisez la classification de `presto-config` configuration pour définir le `graceful-shutdown-timeout` paramètre sur une valeur en secondes ou minutes supérieure à zéro. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

Par exemple, l'augmentation de la valeur `graceful-shutdown-timeout` à "30m" spécifie un délai d'attente de 30 minutes. À la fin du délai d'arrêt, le nœud marqué pour la désaffectation fait l'objet d'un arrêt forcé s'il attend que des tâches de requête se terminent. Dans ce cas, la requête échoue. Si les tâches de requête se terminent dans cinq minutes, le nœud marqué pour la mise hors service s'arrête dans cinq minutes, à condition que les autres YARN applications aient terminé leur exécution.

Exemple Exemple de configuration de dimensionnement automatique Presto avec désaffectation gracieuse

Remplacez la valeur `graceful-shutdown-timeout` par le nombre de minutes approprié pour votre configuration. Il n'y a pas de valeur maximale. L'exemple ci-dessous définit une valeur de délai d'arrêt de 1800 secondes (30 minutes).

```
[
  {
    "classification": "presto-config",
    "properties": {
      "graceful-shutdown-timeout": "1800s"
    }
  }
]
```

Limites

La mise hors service gracieuse de PrestoDB ne fonctionne pas sur les clusters HTTP où la connectivité est désactivée, par exemple quand elle est définie EMR sur `http-server.http.enabled false`. Trino ne prend pas du tout en charge la mise hors service gracieuse, quel que soit le paramètre `http-server.http.enabled`.

Considérations relatives à Presto sur Amazon EMR

Tenez compte des limites suivantes lorsque vous exécutez [Presto](#) sur AmazonEMR.

Exécutable en ligne de commande Presto

Dans AmazonEMR, PrestoDB et Trino utilisent tous deux le même exécutable en ligne de commande `presto-cli`, comme dans l'exemple suivant.

```
presto-cli --catalog hive
```

Propriétés de déploiement Presto non configurables

La version d'Amazon EMR que vous utilisez détermine les configurations de déploiement Presto disponibles. Pour plus d'informations sur ces propriétés de configuration, consultez la section [Déploiement de Presto](#) dans la documentation de Presto. Le tableau suivant montre les différentes options de configuration pour les fichiers Presto `properties`.

Fichier	Configurable
<code>log.properties</code>	<p>PrestoDB : configurable dans les versions 4.0.0 et ultérieures d'EMR Amazon. Utilisez la classification de configuration <code>presto-log</code> .</p> <p>Trino (PrestoSQL) : configurable dans les EMR versions 6.1.0 et ultérieures d'Amazon. Utilisez la classification de configuration <code>prestoql-log</code> ou <code>trino-log</code> .</p>
<code>config.properties</code>	<p>PrestoDB : configurable dans les versions 4.0.0 et ultérieures d'EMR Amazon. Utilisez la classification de configuration <code>presto-config</code> .</p>

Fichier	Configurable
	Trino (PrestoSQL) : configurable dans les EMR versions 6.1.0 et ultérieures d'Amazon. Utilisez la classification de configuration <code>prestoql-config</code> ou <code>trino-config</code> .
<code>hive.properties</code>	<p>PrestoDB : configurable dans les versions 4.1.0 et ultérieures d'EMR Amazon. Utilisez la classification de configuration <code>presto-connector-hive</code> .</p> <p>Trino (PrestoSQL) : configurable dans les EMR versions 6.1.0 et ultérieures d'Amazon. Utilisez la classification de configuration <code>prestoql-connector-hive</code> ou <code>trino-connector-hive</code> .</p>
<code>node.properties</code>	<p>PrestoDB : configurable dans EMR Amazon version 5.6.0 et versions ultérieures. Utilisez la classification de configuration <code>presto-node</code> .</p> <p>Trino (PrestoSQL) : configurable dans les EMR versions 6.1.0 et ultérieures d'Amazon. Utilisez la classification de configuration <code>prestoql-node</code> ou <code>trino-node</code> .</p>
<code>jvm.config</code>	Non configurable.

Installation de PrestoDB et Trino

Le nom de l'application Presto continue d'être utilisé pour installer PrestoDB sur des clusters. Pour installer Trino sur des clusters, utilisez le nom de l'application Trino (ou Presto SQL dans les anciennes versions d'Amazon). EMR

Vous pouvez installer PrestoDB ou Trino, mais vous ne pouvez pas installer les deux sur un seul cluster. Si vous spécifiez à la fois Prestodb et Trino lorsque vous tentez de créer un cluster, une erreur de validation se produit et la demande de création de cluster échoue.

EMRFS configuration de PrestoOS3 FileSystem

Avec EMR les versions 5.12.0 et ultérieures d'Amazon, PrestoDB peut utiliser EMRFS. Il s'agit de la configuration par défaut. EMRFS est également le système de fichiers par défaut pour Trino (PrestoSQL) dans les EMR versions 6.1.0 et ultérieures d'Amazon. Pour plus d'informations, consultez [EMR File System \(EMRFS\)](#) dans le Amazon EMR Management Guide. Dans les versions antérieures d'Amazon EMR, PrestoOS3 FileSystem est la seule option de configuration.

Vous pouvez utiliser une configuration de sécurité pour configurer le chiffrement EMRFS des données dans Amazon S3. Vous pouvez également utiliser IAM des rôles pour les EMRFS demandes adressées à Amazon S3. Pour plus d'informations, consultez [Comprendre les options de chiffrement](#) et [Configurer IAM les rôles pour les EMRFS demandes adressées à Amazon S3](#) dans le guide EMR de gestion Amazon.

Note

Si vous interrogez des données sous-jacentes dans Amazon S3 avec Amazon EMR version 5.12.0, des erreurs Presto peuvent se produire. Cela est dû au fait que Presto ne parvient pas à récupérer les valeurs de classification de configuration depuis `emrfs-site.xml`. Pour contourner le problème, créez un sous-répertoire `emrfs` sous `usr/lib/presto/plugin/hive-hadoop2/` et un lien symbolique dans `usr/lib/presto/plugin/hive-hadoop2/emrfs` vers le fichier `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` existant. Redémarrez ensuite le processus `presto-server` (`sudo presto-server stop` suivi de `sudo presto-server start`).

Vous pouvez remplacer la EMRFS valeur par défaut et utiliser PrestoOS3 à la place FileSystem. Pour ce faire, utilisez la classification de configuration `presto-connector-hive` pour définir `hive.s3-file-system-type` sur PRESTO comme illustré dans l'exemple suivant. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.s3-file-system-type": "PRESTO"
    }
  }
]
```

Si vous utilisez PrestoS3FileSystem, utilisez la classification de `presto-connector-hive` configuration ou utilisez Trino `trino-connector-hive` pour configurer les propriétés de PrestoS3FileSystem. Pour plus d'informations sur les propriétés disponibles, consultez [Configuration Amazon S3](#) dans la section du connecteur Hive de la documentation Presto. Ces paramètres ne s'appliquent pas à EMRFS.

Paramètre par défaut de l'emprunt d'identité de l'utilisateur final

Par défaut, EMR les versions 5.12.0 et ultérieures d'Amazon autorisent l'usurpation d'identité de l'utilisateur final pour accéder à HDFS. Pour plus d'informations, consultez [Emprunt de l'identité de l'utilisateur final](#) dans la documentation Presto. Pour modifier ce paramètre avec la classification de configuration `presto-config`, définissez la propriété `hive.hdfs.impersonation.enabled` sur `false`.

Port par défaut de l'interface Web Presto

Par défaut, Amazon EMR configure l'interface Web Presto sur le coordinateur Presto pour utiliser le port 8889 (pour PrestoDB et Trino). Pour modifier le port, utilisez la classification de configuration `presto-config` pour définir la propriété `http-server.http.port`. Pour plus d'informations, consultez [Propriétés de configuration](#) dans la section Déploiement de Presto de la documentation Presto.

Problème d'exécution du compartiment Hive dans certaines versions

Presto version 152.3 connaît un problème lié à l'exécution de compartiment Hive qui entraîne des performances des requêtes Presto significativement plus lentes dans certaines circonstances. EMR les versions 5.0.3, 5.1.0 et 5.2.0 d'Amazon incluent cette version de Presto. Pour atténuer ce problème, utilisez la classification de configuration `presto-connector-hive` pour définir la propriété `hive.bucket-execution` sur `false` tel qu'illustré dans l'exemple suivant.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.bucket-execution": "false"
    }
  }
]
```

Historique des versions de Presto

Le tableau suivant répertorie la version de Presto incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section [Version des composants de votre version](#) dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Presto

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-7.2.0	0,285	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.36.2	0,267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-7.1.0	0,284	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-7.0.0	0,283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.15.0	0,283	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.14.0	0,281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.13.0	0,281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.12.0	0,281	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.11.1	0,279	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.11.0	0,279	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.10.1	0,278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.10.0	0,278	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.9.1	0,276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.9.0	0,276	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.8.1	0,273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.8.0	0,273	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.7.0	0,272	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.36.1	0,267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.36.0	0,267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.6.0	0,267	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.35.0	0,266	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.5.0	0,261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.4.0	0,254,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.3.1	0,245,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.3.0	0,245,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.2.1	0,238,3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.2.0	0,238,3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.1.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.1.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-6.0.1	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-6.0.0	0.230	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.34.0	0,261	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.33.1	0,245,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.33.0	0,245,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.32.1	0,240,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.32.0	0,240,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.31.1	0,238,3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.31.0	0,238,3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.30.2	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
emr-5.30.1	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.30.0	0.232	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mariadb-server, presto-coordinator, presto-worker
mr-5.29.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.28.1	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.28.0	0.227	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-presto, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.27.1	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.27.0	0.224	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.26.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.25.0	0.220	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.24.1	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.24.0	0.219	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.23.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.23.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.22.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.2	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.21.1	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.21.0	0.215	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.20.1	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.20.0	0.214	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.19.1	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.19.0	0.212	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.18.1	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.18.0	0.210	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.17.2	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.17.1	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.17.0	0.206	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.16.1	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.16.0	0.203	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.15.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.15.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.2	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.14.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.14.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.13.1	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.13.0	0.194	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.12.3	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.2	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.12.1	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.12.0	0.188	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.11.4	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.3	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.11.2	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.11.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.11.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.10.1	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.10.0	0.187	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.9.1	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.9.0	0.184	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.8.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.8.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.8.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.7.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.7.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.6.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.6.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.4	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.3	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.5.2	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.1	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.5.0	0.170	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.4.1	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.4.0	0.166	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.3.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.3.0	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.3	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.2.2	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.1	0.157.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.2.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.1.1	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.1.0	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.3	0.152.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Étiquette Amazon EMR Release	Version Presto	Composants installés avec Presto
emr-5.0.2	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.1	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker
emr-5.0.0	0.150	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, mysql-server, presto-coordinator, presto-worker

Le tableau suivant répertorie la version de Trino (PrestoSQL) incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Presto SQL a changé son nom en Trino à partir de la version 351.

Informations sur la version de Trino (PrestoSQL)

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-7.2.0	436	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-7.1.0	435	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-7.0.0	426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.15.0	426	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.14.0	422	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.13.0	414	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.12.0	414	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.11.1	410	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.11.0	410	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.10.1	403	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.10.0	403	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.9.1	398	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.9.0	398	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.8.1	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.8.0	388	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.7.0	378	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.6.0	367	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.5.0	360	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker
emr-6.4.0	359	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-trino, hcatalog-server, mariadb-server, trino-coordinator, trino-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.3.1	350	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.3.0	350	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.2.1	343	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.2.0	343	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Étiquette Amazon EMR Release	Version Trino (PrestoSQL)	Composants installés avec Trino (PrestoSQL)
emr-6.1.1	338	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker
emr-6.1.0	338	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hive-client, hudi, hudi-prestosql, hcatalog-server, mariadb-server, prestosql-coordinator, prestosql-worker

Notes de mise à jour de Trino (PrestoSQL) par version

- [Amazon EMR 6.9.0 - Notes de mise à jour de Trino \(PrestoSQL\)](#)

Amazon EMR 6.9.0 - Notes de mise à jour de Trino (PrestoSQL)

Amazon EMR 6.9.0 - Nouvelles fonctionnalités de Trino (PrestoSQL)

- Pour prendre en charge les requêtes de longue durée, Trino inclut désormais un mécanisme d'exécution tolérant aux pannes. L'exécution tolérante aux pannes atténue les échecs des requêtes en réessayant les requêtes qui ont échoué ou les tâches correspondantes. Pour de plus amples informations, veuillez consulter [Exécution tolérante aux pannes dans Trino](#).

Amazon EMR 6.9.0 - Modifications apportées à Trino (PrestoSQL)

Amazon EMR 6.9.0 - Modifications apportées à PrestoDB

Type	Description
Upgrade	Mise à niveau de Prestodb vers la version 0.276
Upgrade	Prise en charge pour Hadoop 3.3.3.
Upgrade	Mise à niveau de Hudi vers la version 0.12.1
Fonctionnalité	Intégration à Amazon EMR et Presto AWS Lake Formation pour les charges de travail interactives utilisant. GCSC API
Fonctionnalité	Ajout de configurations liées à Kerberos dans Security Configs for PrestoDB pour activer Keberos.
Correctif de bogue	Inversion de la OSS pull request #18115 qui avait été ajoutée pour réduire le nombre de <code>hdfsConfiguration</code> copies. Cela provoquait une copie de HDFS configuration incorrecte lors de l'utilisation de tables Hudi EMRFS ou Hudi.

Amazon EMR 6.9.0 - Modifications apportées à Trino

Type	Description
Upgrade	Mise à niveau de Trino vers la version 398
Upgrade	Prise en charge pour Hadoop 3.3.3.
Fonctionnalité	Support tardif : ajout de la prise en charge du spouillage d'Exchange sur HDFS Amazon S3. Pour plus d'informations, consultez Exécution tolérante aux pannes dans Trino .
Correctif de bogue	Lorsque Trino Iceberg est utilisé et que le catalogue Glue est activé, évitez d'ajouter un URI de métastore dans <code>iceberg.properties</code> .

Amazon EMR 6.9.0 - Problèmes connus liés à Trino (PrestoSQL)

- Pour la EMR version 6.9.0 d'Amazon, Trino ne fonctionne pas sur les clusters activés pour Apache Ranger. Si vous devez utiliser Trino avec Ranger, contactez [AWS Support](#).

Apache Spark

[Apache Spark](#) est un framework de traitement distribué et un modèle de programmation qui vous aident à effectuer du machine learning, du traitement de flux ou des analyses de graphes avec des EMR clusters Amazon. De manière analogue à Apache Hadoop, Spark est un système de traitement distribué open source, couramment utilisé pour les charges de travail de big data. Cependant, Spark présente plusieurs différences notables par rapport à Hadoop MapReduce. Spark dispose d'un moteur d'exécution de graphes acycliques dirigés (DAG) optimisé et met activement en cache les données en mémoire, ce qui peut améliorer les performances, en particulier pour certains algorithmes et requêtes interactives.

De manière native, Spark prend en charge les applications écrites en Scala, python ou Java. Il inclut également plusieurs bibliothèques étroitement intégrées pour SQL ([Spark SQL](#)), l'apprentissage automatique ([MLlib](#)), le traitement de flux ([Spark streaming](#)) et le traitement de graphes ([GraphX](#)). Ces outils facilitent l'exploitation de l'infrastructure Spark pour une grande variété de cas d'utilisation.

Vous pouvez installer Spark sur un EMR cluster Amazon avec d'autres applications Hadoop, et il peut également exploiter le système de EMR fichiers Amazon (EMRFS) pour accéder directement aux données dans Amazon S3. Hive est également intégré à Spark afin que vous puissiez utiliser un HiveContext objet pour exécuter des scripts Hive à l'aide de Spark. Un contexte Hive est inclus dans le shell de Spark en tant que `sqlContext`.

Pour un exemple de didacticiel sur la configuration d'un EMR cluster avec Spark et l'analyse d'un exemple de jeu de données, consultez [Tutoriel : Getting started with Amazon EMR](#) sur le blog AWS News.

Important

[Apache Spark version 2.3.1, disponible à partir de la EMR version 5.16.0 d'Amazon, adresses CVE-2018-8024 et -2018-1334. CVE](#) Nous vous recommandons de migrer les versions antérieures de Spark vers la version 2.3.1 ou ultérieure.

Le tableau suivant répertorie la version de Spark incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Spark.

Pour connaître la version des composants installés avec Spark dans cette version, consultez la [version 7.2.0 des versions des composants](#).

Informations sur la version de Spark pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version de Spark	Composants installés avec Spark
emr-7.2.0	Spark 3.5.1	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Le tableau suivant répertorie la version de Spark incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Spark.

Pour connaître la version des composants installés avec Spark dans cette version, consultez la rubrique [Versions des composants de la version 6.15.0](#).

Informations sur la version de Spark pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version de Spark	Composants installés avec Spark
emr-6.15.0	Spark 3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-

Étiquette EMR de sortie Amazon	Version de Spark	Composants installés avec Spark
		kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Note

La EMR version 6.8.0 d'Amazon est fournie avec Apache Spark 3.3.0. Cette version de Spark utilise Apache Log4j 2 et le fichier `log4j2.properties` pour configurer Log4j dans les processus Spark. Si vous utilisez Spark dans le cluster ou si vous créez des EMR clusters avec des paramètres de configuration personnalisés, et que vous souhaitez passer à la EMR version 6.8.0 d'Amazon, vous devez migrer vers la nouvelle classification de `spark-log4j2` configuration et le nouveau format de clé pour Apache Log4j 2. Pour de plus amples informations, veuillez consulter [Migration d'Apache Log4j 1.x vers Log4j 2.x](#).

Le tableau suivant répertorie la version de Spark incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Spark.

Pour la version des composants installés avec Spark dans cette version, consultez la [version 5.36.2 Versions des composants](#).

Informations sur la version de Spark pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version de Spark	Composants installés avec Spark
emr-5.36.2	Spark 2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode,

Étiquette EMR de sortie Amazon	Version de Spark	Composants installés avec Spark
		hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Rubriques

- [Création d'un cluster avec Apache Spark](#)
- [Exécutez des applications Spark avec Docker sur Amazon 6.x EMR](#)
- [Utiliser le catalogue de données AWS Glue comme métastore pour Spark SQL](#)
- [Configuration de Spark](#)
- [Optimisation des performances de Spark](#)
- [Mise en cache des fragments de résultats Spark](#)
- [Utiliser l'RAPIDSaccélérateur Nvidia pour Apache Spark](#)
- [Accès au shell de Spark](#)
- [Utiliser Amazon SageMaker Spark pour le machine learning](#)
- [Ecriture d'une application Spark](#)
- [Améliorer les performances de Spark avec Amazon S3](#)
- [Ajout d'une étape Spark](#)
- [Afficher l'historique de l'application Spark](#)
- [Accédez au site Web Spark UIs](#)
- [Utilisation du connecteur Amazon Kinesis Data Streams pour le streaming structuré Spark](#)
- [Utilisation de l'intégration d'Amazon Redshift pour Apache Spark avec Amazon EMR](#)
- [Historique des versions de Spark](#)

Création d'un cluster avec Apache Spark

La procédure suivante crée un cluster avec [Spark](#) installé à l'aide des options rapides de la EMR console Amazon.

Vous pouvez alternativement utiliser Options avancées pour personnaliser davantage la configuration de votre cluster, ou pour soumettre des étapes d'installation par programmation d'applications, puis d'exécution d'applications personnalisées. Quelle que soit l'option de création de cluster, vous pouvez choisir d'utiliser AWS Glue comme SQL métastore Spark. Pour plus d'informations, consultez [Utiliser le catalogue de données AWS Glue comme métastore pour Spark SQL](#).

Pour lancer un cluster avec Spark installé

1. Ouvrez la EMR console Amazon à l'adresse <https://console.aws.amazon.com/emr>.
2. Choisissez Créer un cluster pour utiliser Options rapides.
3. Entrez un nom de cluster. Le nom de votre cluster ne peut pas contenir les caractères <, >, \$, | ou ` (backtick).
4. Pour Configuration logicielle, choisissez une option de Version.
5. Sous Applications, choisissez la solution groupée d'applications Spark.
6. Sélectionnez d'autres options si nécessaire, puis choisissez Create cluster (Créer le cluster).

Note

Pour configurer Spark lorsque vous créez le cluster, consultez [Configuration de Spark](#).

Pour lancer un cluster sur lequel Spark est installé à l'aide du AWS CLI

- Créez le cluster à l'aide de la commande suivante.

```
aws emr create-cluster --name "Spark cluster" --release-label emr-7.2.0 --
applications Name=Spark \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 --use-
default-roles
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

Pour lancer un cluster sur lequel Spark est installé à l'aide de

Spécifiez Spark comme application avec `SupportedProductConfig` utilisé dans `RunJobFlowRequest`.

- L'exemple suivant montre comment créer un cluster avec Spark au moyen de Java.

```
import com.amazonaws.AmazonClientException;
import com.amazonaws.auth.AWSCredentials;
import com.amazonaws.auth.AWSStaticCredentialsProvider;
import com.amazonaws.auth.profile.ProfileCredentialsProvider;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.*;
import com.amazonaws.services.elasticmapreduce.util.StepFactory;

public class Main {

    public static void main(String[] args) {
        AWSCredentials credentials_profile = null;
        try {
            credentials_profile = new
ProfileCredentialsProvider("default").getCredentials();
        } catch (Exception e) {
            throw new AmazonClientException(
                "Cannot load credentials from .aws/
credentials file. " +
                "Make sure that the
credentials file exists and the profile name is specified within it.",
                e);
        }

        AmazonElasticMapReduce emr =
AmazonElasticMapReduceClientBuilder.standard()
```

```
        .withCredentials(new
AWSStaticCredentialsProvider(credentials_profile))
        .withRegion(Regions.US_WEST_1)
        .build();

// create a step to enable debugging in the AWS Management Console
StepFactory stepFactory = new StepFactory();
StepConfig enabledebugging = new StepConfig()
    .withName("Enable debugging")
    .withActionOnFailure("TERMINATE_JOB_FLOW")

.withHadoopJarStep(stepFactory.newEnableDebuggingStep());

Application spark = new Application().withName("Spark");

RunJobFlowRequest request = new RunJobFlowRequest()
    .withName("Spark Cluster")
    .withReleaseLabel("emr-5.20.0")
    .withSteps(enabledebugging)
    .withApplications(spark)
    .withLogUri("s3://path/to/my/logs/")
    .withServiceRole("EMR_DefaultRole")
    .withJobFlowRole("EMR_EC2_DefaultRole")
    .withInstances(new JobFlowInstancesConfig()
        .withEc2SubnetId("subnet-12ab3c45")
        .withEc2KeyName("myEc2Key")
        .withInstanceCount(3)

.withKeepJobFlowAliveWhenNoSteps(true)
        .withMasterInstanceType("m4.large")

.withSlaveInstanceType("m4.large"));
    RunJobFlowResult result = emr.runJobFlow(request);
    System.out.println("The cluster ID is " + result.toString());
}
}
```

Exécutez des applications Spark avec Docker sur Amazon 6.x EMR

Avec Amazon EMR 6.0.0, les applications Spark peuvent utiliser des conteneurs Docker pour définir les dépendances de leurs bibliothèques, au lieu d'installer des dépendances sur les EC2 instances Amazon individuelles du cluster. Pour exécuter Spark avec Docker, vous devez d'abord configurer

le registre Docker et définir des paramètres supplémentaires lors de la soumission d'une application Spark. Pour plus d'informations, consultez [Configuration de l'intégration Docker](#).

Lorsque l'application est soumise, YARN appelle Docker pour extraire l'image Docker spécifiée et exécuter l'application Spark dans un conteneur Docker. Cette approche vous permet de définir et d'isoler facilement les dépendances. Cela réduit le temps de démarrage ou de préparation des instances dans le EMR cluster Amazon avec les bibliothèques nécessaires à l'exécution des tâches.

Considérations lors de l'exécution de Spark avec Docker

Lorsque vous exécutez Spark avec Docker, assurez-vous que les conditions suivantes sont remplies :

- Le `docker` package et les nœuds de tâches ne CLI sont installés que sur les nœuds principaux et les nœuds de tâches.
- Sur Amazon EMR 6.1.0 et versions ultérieures, vous pouvez également installer Docker sur un nœud principal à l'aide des commandes suivantes.

```
sudo yum install -y docker
sudo systemctl start docker
```

- La `spark-submit` commande doit toujours être exécutée à partir d'une instance principale du EMR cluster Amazon.
- Les registres Docker utilisés pour résoudre les images Docker doivent être définis à l'aide de la classification API avec la clé de `container-executor` classification pour définir des paramètres supplémentaires lors du lancement du cluster :
 - `docker.trusted.registries`
 - `docker.privileged-containers.registries`
- Pour exécuter une application Spark dans un conteneur Docker, les options de configuration suivantes sont nécessaires :
 - `YARN_CONTAINER_RUNTIME_TYPE=docker`
 - `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={DOCKER_IMAGE_NAME}`
- Lorsque vous utilisez Amazon ECR pour récupérer des images Docker, vous devez configurer le cluster pour qu'il s'authentifie. Pour ce faire, vous devez utiliser l'option de configuration suivante :
 - `YARN_CONTAINER__RUNTIME__DOCKER__CLIENT__CONFIG = {DOCKER__CLIENT__CONFIG_PATH_HDFS_ON_}`

- Dans Amazon EMR 6.1.0 et versions ultérieures, vous n'êtes pas obligé d'utiliser la commande répertoriée
`YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON_HDFS}` lorsque la fonctionnalité d'authentification ECR automatique est activée.
- Toute image Docker utilisée avec Spark implique l'installation préalable de Java dans l'image Docker.

Pour plus d'informations sur les conditions préalables, consultez [Configuration de l'intégration de Docker](#).

Création d'une image Docker

Les images Docker sont créées à l'aide d'un Dockerfile, qui définit les packages et la configuration à inclure dans l'image. Les deux exemples suivants utilisent Dockerfiles et PySpark SparkR.

PySpark Dockerfile

Les images Docker créées à partir de ce Dockerfile incluent Python 3 et le package Python NumPy . Ce Dockerfile utilise Amazon Linux 2 et Amazon Corretto 8. JDK

```
FROM amazoncorretto:8

RUN yum -y update
RUN yum -y install yum-utils
RUN yum -y groupinstall development

RUN yum list python3*
RUN yum -y install python3 python3-dev python3-pip python3-virtualenv

RUN python -V
RUN python3 -V

ENV PYSPARK_DRIVER_PYTHON python3
ENV PYSPARK_PYTHON python3

RUN pip3 install --upgrade pip
RUN pip3 install numpy pandas

RUN python3 -c "import numpy as np"
```

Dockerfile SparkR

Les images Docker créées à partir de ce Dockerfile incluent R et le package. randomForest CRAN
Ce Dockerfile inclut Amazon Linux 2 et Amazon Corretto 8. JDK

```
FROM amazoncorretto:8

RUN java -version

RUN yum -y update
RUN amazon-linux-extras install R4

RUN yum -y install curl hostname

#setup R configs
RUN echo "r <- getOption('repos'); r['CRAN'] <- 'http://cran.us.r-project.org';
  options(repos = r);" > ~/.Rprofile

RUN Rscript -e "install.packages('randomForest')"
```

Pour plus d'informations sur la syntaxe Dockerfile, consultez la [documentation de référence relative à Dockerfile](#).

Utilisation d'images Docker provenant d'Amazon ECR

Amazon Elastic Container Registry (Amazon ECR) est un registre de conteneurs Docker entièrement géré, qui facilite le stockage, la gestion et le déploiement d'images de conteneurs Docker. Lorsque vous utilisez Amazon ECR, le cluster doit être configuré pour faire confiance à votre instance de ECR, et vous devez configurer l'authentification pour que le cluster utilise les images Docker d'Amazon ECR. Pour plus d'informations, consultez [Configuration YARN pour accéder à Amazon ECR](#).

Pour garantir que les EMR hôtes Amazon peuvent accéder aux images stockées sur Amazon ECR, votre cluster doit disposer des autorisations définies dans la AmazonEC2ContainerRegistryReadOnly politique associée au profil d'instance. Pour de plus amples informations, veuillez consulter [Stratégie AmazonEC2ContainerRegistryReadOnly](#).

Dans cet exemple, le cluster doit être créé avec la configuration supplémentaire suivante pour garantir la fiabilité ECR du registre Amazon. Remplacez le *123456789123.dkr.ecr.us-east-1.amazonaws.com* point de terminaison avec votre point de ECR terminaison Amazon.

```
[
  {
```

```
"Classification": "container-executor",
"Configurations": [
  {
    "Classification": "docker",
    "Properties": {
      "docker.privileged-containers.registries":
"local,centos,123456789123.dkr.ecr.us-east-1.amazonaws.com",
      "docker.trusted.registries": "local,centos,123456789123.dkr.ecr.us-
east-1.amazonaws.com"
    }
  }
],
"Properties": {}
}
```

Utilisation PySpark avec Amazon ECR

L'exemple suivant utilise le PySpark Dockerfile, qui sera balisé et chargé sur Amazon. ECR Après avoir chargé le Dockerfile, vous pouvez exécuter la PySpark tâche et vous référer à l'image Docker d'Amazon. ECR

Après avoir lancé le cluster, connectez-vous SSH à un nœud principal et exécutez les commandes suivantes pour créer l'image Docker locale à partir de l'exemple PySpark Dockerfile.

Tout d'abord, créez un répertoire et un fichier Dockerfile.

```
mkdir pyspark
vi pyspark/Dockerfile
```

Collez le contenu du PySpark Dockerfile et exécutez les commandes suivantes pour créer une image Docker.

```
sudo docker build -t local/pyspark-example pyspark/
```

Créez le `emr-docker-examples` ECR référentiel pour les exemples.

```
aws ecr create-repository --repository-name emr-docker-examples
```

Marquez et téléchargez l'image créée localement vers ECR, en remplaçant

123456789123.dkr.ecr.us-east-1.amazonaws.com avec votre ECR terminal.

```
sudo docker tag local/pyspark-example 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:pyspark-example
sudo docker push 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:pyspark-example
```

SSHUtilisez-le pour vous connecter au nœud principal et préparer un script Python avec le nom du fichier `main.py`. Collez le contenu suivant dans le fichier `main.py` et enregistrez-le.

```
from pyspark.sql import SparkSession
spark = SparkSession.builder.appName("docker-numpy").getOrCreate()
sc = spark.sparkContext

import numpy as np
a = np.arange(15).reshape(3, 5)
print(a)
```

Sur Amazon EMR 6.0.0, pour soumettre la tâche, référez le nom de l'image Docker. Définissez les paramètres de configuration supplémentaires pour vous assurer que l'exécution de la tâche utilise Docker comme environnement d'exécution. Lorsque vous utilisez Amazon ECR, vous devez référencer le fichier `config.json` contenant les informations d'identification utilisées pour s'authentifier auprès d'Amazon ECR.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:pyspark-example
DOCKER_CLIENT_CONFIG=hdfs:///user/hadoop/config.json
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=$DOCKER_CLIENT_CONFIG \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=$DOCKER_CLIENT_CONFIG \
--num-executors 2 \
main.py -v
```

Sur Amazon EMR 6.1.0 et versions ultérieures, pour soumettre la tâche, référez le nom de l'image Docker. Lorsque l'authentification ECR automatique est activée, exécutez la commande suivante.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-  
examples:pyspark-example  
spark-submit --master yarn \  
--deploy-mode cluster \  
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \  
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \  
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \  
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \  
--num-executors 2 \  
main.py -v
```

Lorsque le travail est terminé, prenez note de l'ID de l'YARN application et utilisez la commande suivante pour obtenir le résultat du PySpark travail.

```
yarn logs --applicationId application_id | grep -C2 '\[ \['  
LogLength:55  
LogContents:  
[[ 0  1  2  3  4]  
 [ 5  6  7  8  9]  
[10 11 12 13 14]]
```

Utiliser SparkR avec Amazon ECR

L'exemple suivant utilise le Dockerfile SparkR, qui sera balisé et téléchargé vers ECR. Une fois le Dockerfile chargé, vous pouvez exécuter le job SparkR et vous référer à l'image Docker d'Amazon ECR.

Après avoir lancé le cluster, connectez-vous SSH à un nœud principal et exécutez les commandes suivantes pour créer l'image Docker locale à partir de l'exemple SparkR Dockerfile.

Tout d'abord, créez un répertoire et un fichier Dockerfile.

```
mkdir sparkr  
vi sparkr/Dockerfile
```

Collez le contenu du fichier Dockerfile SparkR et exécutez les commandes suivantes pour créer une image Docker.

```
sudo docker build -t local/sparkr-example sparkr/
```

Marquez et téléchargez l'image créée localement sur Amazon ECR, en remplaçant `123456789123.dkr.ecr.us-east-1.amazonaws.com` avec votre point de ECR terminaison Amazon.

```
sudo docker tag local/sparkr-example 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:sparkr-example
sudo docker push 123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:sparkr-example
```

SSHUtilisez-le pour vous connecter au nœud principal et préparer un script R portant le nom `sparkR.R`. Collez le contenu suivant dans le fichier `sparkR.R`.

```
library(SparkR)
sparkR.session(appName = "R with Spark example", sparkConfig =
  list(spark.some.config.option = "some-value"))

sqlContext <- sparkRSQL.init(spark.sparkContext)
library(randomForest)
# check release notes of randomForest
rfNews()

sparkR.session.stop()
```

Sur Amazon EMR 6.0.0, pour soumettre la tâche, reportez-vous au nom de l'image Docker. Définissez les paramètres de configuration supplémentaires pour vous assurer que l'exécution de la tâche utilise Docker comme environnement d'exécution. Lorsque vous utilisez Amazon ECR, vous devez vous référer au fichier `config.json` contenant les informations d'identification utilisées pour s'authentifier auprès de ECR.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-examples:sparkr-example
DOCKER_CLIENT_CONFIG=hdfs:///user/hadoop/config.json
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=$DOCKER_CLIENT_CONFIG \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
```

```
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG=
$DOCKER_CLIENT_CONFIG \
sparkR.R
```

Sur Amazon EMR 6.1.0 et versions ultérieures, pour soumettre la tâche, référez le nom de l'image Docker. Lorsque l'authentification ECR automatique est activée, exécutez la commande suivante.

```
DOCKER_IMAGE_NAME=123456789123.dkr.ecr.us-east-1.amazonaws.com/emr-docker-
examples:sparkr-example
spark-submit --master yarn \
--deploy-mode cluster \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.executorEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_TYPE=docker \
--conf spark.yarn.appMasterEnv.YARN_CONTAINER_RUNTIME_DOCKER_IMAGE=$DOCKER_IMAGE_NAME \
sparkR.R
```

Lorsque la tâche est terminée, notez l'ID de l'YARNapplication et utilisez la commande suivante pour obtenir le résultat de la tâche SparkR. Cet exemple inclut des tests visant à s'assurer que la randomForest bibliothèque, la version installée et les notes de version sont disponibles.

```
yarn logs --applicationId application_id | grep -B4 -A10 "Type rfNews"
randomForest 4.6-14
Type rfNews() to see new features/changes/bug fixes.
Wishlist (formerly TODO):

* Implement the new scheme of handling classwt in classification.

* Use more compact storage of proximity matrix.

* Allow case weights by using the weights in sampling?

=====
Changes in 4.6-14:
```

Utiliser le catalogue de données AWS Glue comme métastore pour Spark SQL

À l'aide de la EMR version 5.8.0 ou ultérieure d'Amazon, vous pouvez configurer Spark SQL pour utiliser le catalogue de données AWS Glue comme métastore. Cette configuration est recommandée

si vous avez besoin d'un métastore persistant ou d'un métastore partagé par différents clusters, services, applications ou comptes AWS .

AWS Glue est un service d'extraction, de transformation et de chargement (ETL) entièrement géré qui permet de classer vos données de manière simple et rentable, de les nettoyer, de les enrichir et de les déplacer de manière fiable entre différents magasins de données. Le catalogue de données AWS Glue fournit un référentiel de métadonnées unifié couvrant une variété de sources de données et de formats de données, s'intégrant à Amazon EMR ainsi qu'à AmazonRDS, Amazon Redshift, Redshift Spectrum, Athena et à toute application compatible avec le métastore Apache Hive. AWS Les robots d'exploration Glue peuvent automatiquement déduire un schéma à partir des données sources dans Amazon S3 et stocker les métadonnées associées dans le catalogue de données. Pour plus d'informations sur le catalogue de données, consultez la section [Remplissage du catalogue de données AWS Glue](#) dans le AWS Glue Developer Guide.

Des frais distincts s'appliquent pour AWS Glue. Il existe un tarif mensuel pour le stockage et l'accès aux métadonnées du catalogue de données, un taux horaire facturé par minute pour les ETL tâches AWS Glue et le temps d'exécution du robot d'exploration, et un taux horaire facturé par minute pour chaque point de terminaison de développement provisionné. Le catalogue de données vous permet de stocker jusqu'à un million d'objets sans aucuns frais. Si vous stockez plus d'un million d'objets, vous êtes facturé USD 1\$ pour 100 000 objets de plus d'un million. Dans le catalogue de données, sont considérés comme des objets une table, une partition ou une base de données. Pour de plus amples informations, consultez [Tarification Glue](#).

Important

Si vous avez créé des tables à l'aide d'Amazon Athena ou d'Amazon Redshift Spectrum avant le 14 août 2017, les bases de données et les tables sont stockées dans un catalogue géré par Athena, distinct du catalogue de données Glue. AWS Pour intégrer Amazon EMR à ces tables, vous devez passer au catalogue de données AWS Glue. Pour plus d'informations, consultez la section [Mise à niveau vers le catalogue de données AWS Glue](#) dans le guide de l'utilisateur Amazon Athena.

Spécifier le catalogue de données AWS Glue comme métastore

Vous pouvez spécifier le catalogue de données AWS Glue comme métastore à l'aide de AWS Management Console AWS CLI, ou Amazon. EMR API Lorsque vous utilisez le CLI ou API, vous utilisez la classification de configuration de Spark pour spécifier le catalogue de données. En

outre, avec Amazon EMR 5.16.0 et versions ultérieures, vous pouvez utiliser la classification de configuration pour spécifier un catalogue de données dans un autre. Compte AWS Lorsque vous utilisez la console, vous pouvez spécifier le catalogue de données à l'aide des Options avancées ou des Options rapides.

Note

La possibilité d'utiliser AWS Glue Data Catalog est également disponible avec Zeppelin car Zeppelin est installé avec des composants Spark. SQL

Console

Pour spécifier AWS Glue Data Catalog comme métastore Spark avec la nouvelle console

1. Connectez-vous au et ouvrez AWS Management Console la EMR console Amazon à l'adresse <https://console.aws.amazon.com/emr>.
2. Sous Amazon EMR activé EC2 dans le volet de navigation de gauche, choisissez Clusters, puis Create cluster.
3. Sous Solution groupée d'application, choisissez Spark ou Custom. Si vous personnalisez votre cluster, assurez-vous de sélectionner Zeppelin ou Spark comme l'une de vos applications.
4. Sous Paramètres du catalogue de données AWS Glue, cochez la case Utiliser pour les métadonnées de table Spark.
5. Choisissez toutes les autres options qui s'appliquent à votre cluster.
6. Pour lancer cluster, choisissez Créer un cluster.

AWS CLI

Pour spécifier le catalogue de données AWS Glue comme métastore Spark avec le AWS CLI

Pour plus d'informations sur la spécification d'une classification de configuration à l'aide de AWS CLI et Amazon EMRAPI, consultez [Configuration des applications](#).

- Spécifiez la valeur de `hive.metastore.client.factory.class` à l'aide de la classification `spark-hive-site` comme illustré dans l'exemple suivant :

```
[
```



```
{
  "Classification": "spark-hive-site",
  "Properties": {
    "hive.metastore.client.factory.class":
    "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
  }
}
```

Pour spécifier un catalogue de données dans un autre AWS compte, ajoutez la `hive.metastore.glue.catalogid` propriété comme indiqué dans l'exemple suivant. Remplacez *acct-id* par le compte AWS du catalogue de données.

```
[
  {
    "Classification": "spark-hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

IAM autorisations

Le profil d'EC2 instance d'un cluster doit disposer d'IAM autorisations pour les actions AWS Glue. En outre, si vous activez le chiffrement pour les objets AWS Glue Data Catalog, le rôle doit également être autorisé à chiffrer, déchiffrer et générer les objets AWS KMS key utilisés pour le chiffrement.

Autorisations pour les actions AWS Glue

Si vous utilisez le profil d'EC2 instance par défaut pour Amazon EMR, aucune action n'est requise. La politique `AmazonElasticMapReduceforEC2Role` gérée attachée au `EMR_EC2_DefaultRole` autorise toutes les actions AWS Glue nécessaires. Toutefois, si vous spécifiez un profil d'EC2 instance personnalisé et des autorisations, vous devez configurer les actions AWS Glue appropriées. Utilisez la stratégie gérée `AmazonElasticMapReduceforEC2Role` comme point

de départ. Pour plus d'informations, consultez la section [Rôle de service pour les EC2 instances de cluster \(profil d'EC2instance\)](#) dans le guide EMR de gestion Amazon.

Autorisations pour chiffrer et déchiffrer le catalogue de données Glue AWS

Votre profil d'instance a besoin d'autorisations pour chiffrer et déchiffrer les données au moyen de votre clé. Vous n'avez pas besoin de configurer ces autorisations si les deux déclarations suivantes s'appliquent :

- Vous activez le chiffrement des objets du catalogue de données AWS Glue à l'aide de clés gérées pour AWS Glue.
- Vous utilisez un cluster Compte AWS identique au AWS Glue Data Catalog.

Dans le cas contraire, vous devez ajouter l'instruction suivante à la politique d'autorisation attachée à votre profil d'EC2instance.

```
[
  {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Action": [
          "kms:Decrypt",
          "kms:Encrypt",
          "kms:GenerateDataKey"
        ],
        "Resource": "arn:aws:kms:region:acct-
id:key/12345678-1234-1234-1234-123456789012"
      }
    ]
  }
]
```

Pour plus d'informations sur le chiffrement du catalogue de données AWS Glue, consultez la section [Chiffrer votre catalogue de données](#) dans le guide du développeur de AWS Glue.

Autorisations basées sur les ressources

Si vous utilisez AWS Glue conjointement avec Hive, Spark ou Presto sur Amazon, EMR AWS Glue prend en charge les politiques basées sur les ressources afin de contrôler l'accès aux ressources

du catalogue de données. Ces ressources comprennent les bases de données, les tables, les connexions et les fonctions définies par l'utilisateur. Pour plus d'informations, consultez [Politiques de ressources AWS Glue](#) dans le Guide du développeur AWS Glue.

Lorsque vous utilisez des politiques basées sur les ressources pour limiter l'accès à AWS Glue depuis AmazonEMR, le principal que vous spécifiez dans la politique d'autorisation doit être le rôle ARN associé au profil d'EC2instance spécifié lors de la création d'un cluster. Par exemple, pour une politique basée sur les ressources attachée à un catalogue, vous pouvez spécifier le rôle du rôle de service par défaut ARN pour les instances de clusterEC2, *EMR_EC2_DefaultRole* comme lePrincipal, en utilisant le format illustré dans l'exemple suivant :

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

Le *acct-id* peut être différent de l'identifiant du compte AWS Glue. Cela permet d'accéder à partir de EMR clusters appartenant à différents comptes. Vous pouvez spécifier plusieurs principaux, chacun provenant d'un compte différent.

Considérations relatives à l'utilisation du catalogue de données AWS Glue

Tenez compte des éléments suivants lorsque vous utilisez AWS Glue Data Catalog comme métastore avec Spark :

- Le fait d'avoir une base de données par défaut sans emplacement URI entraîne des échecs lors de la création d'une table. Pour contourner ce problème, utilisez la clause LOCATION pour spécifier un emplacement de bucket, tel que `s3://EXAMPLE-DOC-BUCKET`, lorsque vous utilisez CREATE TABLE. Vous pouvez aussi créer des tables dans une autre base de données que celle par défaut.
- Il n'est pas possible de renommer des tables depuis AWS Glue.
- Lorsque vous créez une table Hive sans spécifier de LOCATION, les données de la table sont stockées à l'emplacement spécifié par la propriété `hive.metastore.warehouse.dir`. Par défaut, il s'agit d'un emplacement dansHDFS. Si un autre cluster doit accéder à la table, il échoue, sauf s'il dispose d'autorisations adéquates sur le cluster qui a créé la table. En outre, le HDFS stockage étant transitoire, si le cluster se termine, les données de la table sont perdues et la table doit être recréée. Nous vous recommandons de spécifier un LOCATION dans Amazon S3 lorsque vous créez une table Hive à l'aide de AWS Glue. Vous pouvez également utiliser la classification de configuration `hive-site` pour spécifier un emplacement dans Amazon S3 pour `hive.metastore.warehouse.dir`, qui s'applique à toutes les tables Hive. Si une table est créée dans un HDFS emplacement et que le cluster qui l'a créée est toujours en cours d'exécution, vous pouvez mettre à jour l'emplacement de la table vers Amazon S3 depuis AWS Glue. Pour

plus d'informations, consultez [Working with Tables on the AWS Glue Console](#) dans le AWS Glue Developer Guide.

- Les valeurs de partition contenant des guillemets et des apostrophes ne sont pas prises en charge, par exemple `PARTITION (owner="Doe 's")`..
- Les [statistiques sur les colonnes](#) sont prises en charge à partir de la version emr-5.31.0.
- L'utilisation de l'[autorisation Hive](#) n'est pas prise en charge. Comme alternative, envisagez d'utiliser les [politiques basées sur les ressources AWS Glue](#). Pour plus d'informations, consultez [Utiliser des politiques basées sur les ressources pour le catalogue de données Amazon EMR Access to AWS Glue](#).

Configuration de Spark

Vous pouvez configurer [Spark sur Amazon à l'EMR](#) aide de classifications de configuration. Pour plus d'informations sur les classifications de configuration, consultez [Configuration des applications](#).

Les classifications de configuration pour Spark on Amazon EMR sont les suivantes :

- **spark** – Définit la propriété `maximizeResourceAllocation` sur la valeur `true` ou `false`. Lorsque c'est vrai, Amazon configure EMR automatiquement les `spark-defaults` propriétés en fonction de la configuration matérielle du cluster. Pour de plus amples informations, veuillez consulter [Utiliser maximizeResourceAllocation](#).
- **spark-defaults** – Définit les valeurs dans le fichier `spark-defaults.conf`. Pour plus d'informations, consultez [Configuration Spark](#) dans la documentation Spark.
- **spark-env** – Définit les valeurs dans le fichier `spark-env.sh`. Pour plus d'informations, consultez [Variables d'environnement](#) dans la documentation Spark.
- **spark-hive-site** – Définit les valeurs dans le `hive-site.xml` pour Spark.
- **spark-log4j**— (Amazon EMR versions 6.7.x et antérieures) Définit les valeurs dans le `log4j.properties` fichier. Pour plus d'informations, consultez le fichier [log4j.properties.template](#) sur Github.
- **spark-log4j2**— (Amazon EMR versions 6.8.0 et supérieures) Définit les valeurs dans le `log4j2.properties` fichier. Pour plus d'informations, consultez le fichier [log4j2.properties.template](#) sur Github.
- **spark-metrics** – Définit les valeurs dans le fichier `metrics.properties`. Pour les paramètres et plus d'informations, consultez le fichier [metrics.properties.template](#) sur Github, et [Métriques](#) dans la documentation Spark.

Note

Si vous migrez des charges de travail Spark vers Amazon EMR depuis une autre plateforme, nous vous recommandons de tester vos charges de travail avec le [Paramètres par défaut de Spark définis par Amazon EMR](#) avant d'ajouter des configurations personnalisées. La plupart des clients constatent une amélioration des performances grâce à nos paramètres par défaut.

Rubriques


- [Paramètres par défaut de Spark définis par Amazon EMR](#)
- [Configuration de la collecte des déchets Spark sur Amazon EMR 6.1.0](#)
- [Utiliser maximizeResourceAllocation](#)
- [Configuration du comportement de mise hors service du nœud](#)
- [Variable d' ThriftServer environnement Spark](#)
- [Modification des paramètres Spark par défaut](#)
- [Migration d'Apache Log4j 1.x vers Log4j 2.x](#)

Paramètres par défaut de Spark définis par Amazon EMR


Le tableau suivant montre comment Amazon EMR définit les valeurs par défaut `spark-defaults` qui affectent les applications.

Paramètres par défaut de Spark définis par Amazon EMR

Paramètre	Description	Valeur par défaut
<code>spark.executor.memory</code>	Quantité de mémoire à utiliser par processus d'exécution. Par exemple, 1g, 2g.	Ce paramètre est déterminé par les types d'instance de noyau et de tâche dans le cluster.
<code>spark.executor.cores</code>	Le nombre de cœurs à utiliser sur chaque exécuteur.	Ce paramètre est déterminé par les types d'instance de noyau et de tâche dans le cluster.

Paramètre	Description	Valeur par défaut
<code>spark.dynamicAllocation.enabled</code>	Lorsque cela est vrai, utilisez l'allocation dynamique des ressources pour augmenter ou réduire le nombre d'exécuteurs enregistrés avec une application en fonction de la charge de travail.	<code>true</code> (avec Amazon EMR 4.4.0 et versions ultérieures) <div data-bbox="1068 352 1510 667"><p> Note Le service Spark Shuffle est automatiquement configuré par Amazon. EMR</p></div>
<code>spark.sql.hive.advancedPartitionPredicatePushdown.enabled</code>	Lorsque c'est vrai, le transfert avancé des prédicats de partition vers le métastore Hive est activé.	<code>true</code>

Paramètre	Description	Valeur par défaut
<code>spark.sql.hive.stringLikePartitionPredicatePushdown.enabled</code>	Pousse vers le bas les filtres <code>startsWith</code> , <code>contains</code> et <code>endsWith</code> dans le métastore Hive.	<code>true</code>

 **Note**

Glue ne prend pas en charge le push down des prédicats pour `startsWith` , `contains`, ou `endsWith`. Si vous utilisez le métastore Glue et que vous rencontrez des erreurs en raison de l'activation des prédicats pour ces fonctions , définissez cette configuration sur `false`.

Configuration de la collecte des déchets Spark sur Amazon EMR 6.1.0

La définition de configurations personnalisées de collecte des déchets avec `spark.driver.extraJavaOptions` et `spark.executor.extraJavaOptions` entraîne un échec du lancement du pilote ou de l'exécuteur avec Amazon EMR 6.1 en raison d'une configuration de collecte des déchets conflictuelle avec Amazon EMR 6.1.0. Pour Amazon EMR 6.1.0, la configuration de collecte des déchets par défaut est définie via `spark.driver.defaultJavaOptions` et `spark.executor.defaultJavaOptions`. Cette configuration s'applique uniquement à Amazon EMR 6.1.0. JVMles options non liées à la collecte des déchets, telles que celles permettant de configurer logging (`-verbose: class`), peuvent toujours

être définies `extraJavaOptions`. Pour plus d'informations, consultez [Propriétés de l'application Spark](#).

Utiliser `maximizeResourceAllocation`

Pour configurer vos exécuteurs afin qu'ils utilisent le maximum de ressources possibles sur chaque nœud d'un cluster, définissez `maximizeResourceAllocation` sur `true` dans votre classification de configuration spark. `maximizeResourceAllocation` est spécifique à Amazon EMR. Lorsque vous l'activez `maximizeResourceAllocation`, Amazon EMR calcule les ressources de calcul et de mémoire maximales disponibles pour un exécuteur sur une instance du groupe d'instances principal. Il définit ensuite les paramètres correspondants `spark-defaults` en fonction des valeurs maximales calculées.

Amazon EMR calcule les ressources de calcul et de mémoire maximales disponibles pour un exécuteur en fonction d'un type d'instance issu du parc d'instances principal. Étant donné que chaque parc d'instances peut avoir des types et des tailles d'instances différents au sein d'un même parc, la configuration d'exécuteur EMR utilisée par Amazon n'est peut-être pas la meilleure pour vos clusters. Nous vous déconseillons donc d'utiliser les paramètres par défaut lors de l'utilisation d'une allocation de ressources maximale. Configurez des paramètres personnalisés pour vos clusters de flotte d'instances.

Note

Vous ne devez pas utiliser l'option `maximizeResourceAllocation` sur des clusters contenant d'autres applications distribuées telles que HBase. Amazon EMR utilise des YARN configurations personnalisées pour les applications distribuées, qui peuvent entrer en conflit avec `maximizeResourceAllocation` les applications Spark et entraîner leur échec.

Voici un exemple de classification de configuration Spark avec `maximizeResourceAllocation` défini sur `true`.

```
[
  {
    "Classification": "spark",
    "Properties": {
      "maximizeResourceAllocation": "true"
    }
  }
]
```


]

Paramètres configurés dans **spark-defaults** lorsque **maximizeResourceAllocation** est activé

Paramètre	Description	Valeur
spark.default.parallelism	Le nombre de partitions par défaut RDDs est renvoyé par des transformations telles que jointure et parallélisation lorsqu'elles ne sont pas définies par l'utilisateur. <code>reduceByKey</code>	2 fois le nombre de CPU cœurs disponibles pour les YARN conteneurs.
spark.driver.memory	Quantité de mémoire à utiliser pour le processus du pilote, c'est-à-dire où SparkContext est initialisé. (par exemple, 1 g, 2 g).	Le paramétrage est configuré en fonction des types d'instances dans le cluster. Cependant, étant donné que l'application du pilote Spark peut s'exécuter sur l'instance principale ou sur l'une des instances principales (par exemple, en mode YARN client et en mode cluster, respectivement), elle est définie en fonction du plus petit des types d'instances de ces deux groupes d'instances.
spark.executor.memory	Volume de mémoire à utiliser par les processus de l'exécuter. (par exemple, 1g, 2g)	Le paramétrage est configuré en fonction des types d'instances de noyau et de tâches dans le cluster.
spark.executor.cores	Le nombre de cœurs à utiliser sur chaque exécuteur.	Le paramétrage est configuré en fonction des types d'instances de noyau et de tâches dans le cluster.

Paramètre	Description	Valeur
<code>spark.executor.instances</code>	Le nombre d'exécuteurs.	Le paramétrage est configuré en fonction des types d'instances de noyau et de tâches dans le cluster. Définissez, sauf si <code>spark.dynamicAllocation.enabled</code> définit explicitement sur <code>true</code> en même temps.

Configuration du comportement de mise hors service du nœud

Avec les EMR versions 5.9.0 et supérieures d'Amazon, Spark on Amazon EMR inclut un ensemble de fonctionnalités permettant de garantir que Spark gère correctement la terminaison des nœuds en raison d'un redimensionnement manuel ou d'une demande de politique de dimensionnement automatique. Amazon EMR implémente un mécanisme de refus d'inscription dans Spark qui repose sur le mécanisme de YARN mise hors service. Ce mécanisme aide à garantir qu'aucune nouvelle tâche n'est planifiée sur un nœud qui est mis hors service, tout en autorisant dans le même temps que les tâches déjà en cours d'exécution se terminent. En outre, il existe des fonctionnalités qui permettent de rétablir les tâches Spark plus rapidement en cas de perte de blocs aléatoires lors de la terminaison d'un nœud. Le processus de recalcul est déclenché plus tôt et optimisé pour un recalcul plus rapide avec moins de tentatives d'étape ; en outre, il est possible d'empêcher l'échec des tâches dû aux défaillances d'extraction provoquées par les blocs aléatoires manquants.

Important

Le `spark.decommissioning.timeout.threshold` paramètre a été ajouté dans la EMR version 5.11.0 d'Amazon pour améliorer la résilience de Spark lorsque vous utilisez des instances Spot. Dans les versions précédentes, lorsqu'un nœud utilise une instance Spot et que l'instance est terminée en raison du prix de l'offre, Spark peut ne pas gérer correctement la terminaison. Les tâches peuvent échouer et les recalculs aléatoires peuvent nécessiter un temps important. Pour cette raison, nous vous conseillons d'utiliser la version 5.11.0 ou une version ultérieure si vous utilisez des instances Spot.

Paramètres de mise hors service d'un nœud Spark

Paramètre	Description	Valeur par défaut
<code>spark.blacklist.decommissioning.enabled</code>	Lorsqu'il est défini sur <code>true</code> , Spark deny répertorie les nœuds dont l' <code>decommissioning</code> état est dans YARN. Spark ne planifie pas de nouvelles tâches sur les exécuteurs s'exécutant sur ce nœud. Les tâches déjà en cours d'exécution sont autorisées à se terminer.	<code>true</code>
<code>spark.blacklist.decommissioning.timeout</code>	La durée pendant laquelle un nœud dans l'état <code>decommissioning</code> est sur la liste de refus. Par défaut, cette valeur est définie sur une heure, qui est aussi l'heure par défaut pour <code>yarn.resourcemanager.decommissioning.timeout</code> . Pour s'assurer qu'un nœud est placé sur liste noire pendant toute la période de sa mise hors service, définissez une valeur égale ou supérieure à <code>yarn.resourcemanager.decommissioning.timeout</code> . Une fois le délai de mise hors service expiré, le nœud passe à un <code>decommissioned</code> état et Amazon EMR peut mettre fin à l'instance du nœud. EC2 Si des tâches continuent à	1h

Paramètre	Description	Valeur par défaut
	s'exécuter après l'expiration du délai, elles sont perdues ou supprimées, puis replanifiées sur les exécuteurs s'exécutant sur d'autres nœuds.	
<code>spark.decommissioning.timeout.threshold</code>	Disponible dans la EMR version 5.11.0 ou ultérieure d'Amazon. Indiqué en secondes. Lorsqu'un nœud passe à l'état de mise hors service, si l'hôte le met hors service dans un délai égal ou inférieur à cette valeur, Amazon refuse EMR non seulement d'inscrire le nœud, mais nettoie également l'état hôte (comme spécifié par <code>spark.resourceManager.cleanupExpiredHost</code>) sans attendre que le nœud passe à l'état hors service. Cela permet à Spark de mieux traiter les arrêts d'instances Spot car les instances Spot sont mises hors services dans un délai de 20 secondes, et ce quelle que soit la valeur de <code>yarn.resourcemanager.decommissioning.timeout</code> , qui peut ne pas fournir aux autres nœuds suffisamment de temps pour lire les fichiers de lecture aléatoire.	20s

Paramètre	Description	Valeur par défaut
<code>spark.resourceManager.cleanupExpiredHost</code>	Lorsque la valeur est <code>true</code> , Spark annule l'inscription de l'ensemble des données mises en cache et des blocs aléatoires stockés dans les exécuteurs des nœuds qui se trouvent à l'état <code>decommissioned</code> . Le processus de récupération s'en trouve accéléré.	<code>true</code>
<code>spark.stage.attempt.ignoreOnDecommissionFetchFailure</code>	Lorsque la valeur est <code>true</code> , aide à empêcher que les phases Spark n'échouent et n'entraînent l'échec de la tâche en raison du trop grand nombre d'extractions à partir des nœuds mis hors service ayant elles-mêmes échoué. Les extractions défectueuses de blocs aléatoires d'un nœud ayant l'état <code>decommissioned</code> ne sont pas comptabilisées dans le nombre maximal d'extractions défectueuses consécutives.	<code>true</code>

Variable d' ThriftServer environnement Spark

Spark définit la variable d'environnement du port de serveur Thrift Hive, `HIVE_SERVER2_THRIFT_PORT`, sur 10001.

Modification des paramètres Spark par défaut

Vous modifiez les paramètres par défaut dans `spark-defaults.conf` à l'aide de la classification de configuration `spark-defaults` ou du paramètre `maximizeResourceAllocation` dans la classification de configuration `spark`.

Les procédures suivantes indiquent comment modifier les paramètres à l'aide de la console CLI or.

Pour créer un cluster avec `spark.executor.memory` défini sur 2g à l'aide du CLI

- Créez un cluster avec Spark installé et la valeur `spark.executor.memory` définie sur 2g, à l'aide de la commande suivante, qui fait référence à un fichier, `myConfig.json` stocké dans Amazon S3.

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Spark \  
--instance-type m5.xlarge --instance-count 2 --service-role EMR_DefaultRole_V2 \  
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole --configurations https://  
s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

`myConfig.json`:

```
[  
  {  
    "Classification": "spark-defaults",  
    "Properties": {  
      "spark.executor.memory": "2G"  
    }  
  }  
]
```

Pour créer un cluster avec `spark.executor.memory` défini sur 2g à l'aide de la console

1. Accédez à la nouvelle EMR console Amazon et sélectionnez Passer à l'ancienne console dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).
2. Choisissez Créer un cluster et Go to advanced options (Aller aux options avancées).
3. Choisissez Spark.
4. Sous Edit software settings (Modifier les paramètres logiciels), conservez l'option Enter configuration (Saisir la configuration) et saisissez la configuration suivante :

```
classification=spark-defaults,properties=[spark.executor.memory=2G]
```

5. Sélectionnez d'autres options, choisissez , puis Create cluster (Créer un cluster).

À régler `maximizeResourceAllocation`

- Créez un cluster sur lequel Spark est installé et `maximizeResourceAllocation` défini sur `true` en utilisant le AWS CLI, en faisant référence à un fichier `myConfig.json`, stocké dans Amazon S3.

```
aws emr create-cluster --release-label emr-7.2.0 --applications Name=Spark \  
--instance-type m5.xlarge --instance-count 2 --service-role EMR_DefaultRole_V2 \  
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole --configurations https://  
s3.amazonaws.com/mybucket/myfolder/myConfig.json
```

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

`myConfig.json`:

```
[  
  {  
    "Classification": "spark",  
    "Properties": {
```

```
    "maximizeResourceAllocation": "true"  
  }  
}  
]
```

Note

Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier des classifications de configuration supplémentaires pour chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, utilisez la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).

Migration d'Apache Log4j 1.x vers Log4j 2.x

Les versions 3.2.x et antérieures d'[Apache Spark](#) utilisent l'ancien Apache Log4j 1.x et le fichier `log4j.properties` pour configurer Log4j dans les processus Spark. Les versions 3.3.0 et ultérieures d'Apache Spark utilisent Apache Log4j 2.x et le fichier `log4j2.properties` pour configurer Log4j dans les processus Spark.

Si vous avez configuré Apache Spark Log4j à l'aide d'une EMR version d'Amazon inférieure à la version 6.8.0, vous devez supprimer l'ancienne classification de `spark-log4j` configuration et migrer vers la classification de `spark-log4j2` configuration et le format de clé avant de pouvoir passer à Amazon EMR 6.8.0 ou version ultérieure. L'ancienne `spark-log4j` classification entraîne l'échec de la création de clusters avec une `ValidationException` erreur dans les EMR versions 6.8.0 et ultérieures d'Amazon. Aucun frais ne vous sera facturé en cas de panne liée à l'incompatibilité de Log4j, mais vous devez supprimer la classification de configuration obsolète `spark-log4j` pour continuer.

Pour plus d'informations sur la migration d'Apache Log4j 1.x vers Log4j 2.x, consultez le [Guide de migration d'Apache Log4j](#) et le [Modèle Spark Log4j 2](#) sur Github.

Note

Avec AmazonEMR, Apache Spark utilise un `log4j2.properties` fichier plutôt que le fichier `.xml` décrit dans le guide de migration d'[Apache Log4j](#). De plus, nous ne recommandons pas d'utiliser la méthode du pont Log4j 1.x pour convertir en Log4j 2.x.

Optimisation des performances de Spark

Amazon EMR propose plusieurs fonctionnalités d'optimisation des performances pour Spark. Cette rubrique explique en détail chaque fonction d'optimisation.

Pour de plus amples informations sur la définition de la configuration Spark, veuillez consulter [Configuration de Spark](#).

Exécution de requêtes adaptative

L'exécution adaptative des requêtes est un framework permettant de réoptimiser les plans de requêtes en fonction des statistiques d'exécution. À partir d'Amazon EMR 5.30.0, les optimisations d'exécution adaptative des requêtes suivantes d'Apache Spark 3 sont disponibles sur Apache Amazon EMR Runtime pour Spark 2.

- Conversion de jointure adaptative
- Coalescence adaptative de partitions shuffle

Conversion de jointure adaptative

La conversion adaptative des jointures améliore les performances des requêtes en convertissant les `sort-merge-join broadcast-hash-joins` opérations en opérations basées sur la taille d'exécution des étapes de requête. B a `roadcast-hash-joins` tendance à être plus performant lorsqu'un côté de la jointure est suffisamment petit pour diffuser efficacement sa sortie sur tous les exécuteurs, évitant ainsi d'avoir à mélanger les échanges et à trier les deux côtés de la jointure. La conversion adaptative des jointures élargit l'éventail des cas dans lesquels Spark s'exécute `broadcast-hash-joins` automatiquement.

Cette caractéristique est activée par défaut. Il peut être désactivé en le définissant `spark.sql.adaptive.enabled` sur `false`, ce qui désactive également le cadre d'exécution adaptative des requêtes. Spark décide de `sort-merge-join` convertir a en a `broadcast-hash-`

join lorsque la statistique de taille d'exécution de l'un des côtés de la jointure ne dépasse `spark.sql.autoBroadcastJoinThreshold`, la valeur par défaut étant de 10 485 760 octets (10 MiB).

Coalescence adaptative de partitions réorganisées

La fusion adaptative des partitions de distribution améliore les performances des requêtes en fusionnant de petites partitions de distribution contiguës afin d'éviter la surcharge liée à un trop grand nombre de petites tâches. Cela vous permet de configurer un plus grand nombre de partitions de shuffle initiales à l'avance, puis de les réduire à une taille ciblée lors de l'exécution, ce qui augmente les chances d'avoir des partitions shuffle réparties de manière plus uniforme.

Cette fonctionnalité est activée par défaut, sauf si `spark.sql.shuffle.partitions` est explicitement défini. Il peut être activé en réglant `spark.sql.adaptive.coalescePartitions.enabled` sur `true`. Le nombre initial de partitions réorganisées et la taille de la partition cible peuvent être ajustés à l'aide des propriétés `spark.sql.adaptive.coalescePartitions.minPartitionNum` et `spark.sql.adaptive.advisoryPartitionSizeInBytes` respectivement. Consultez le tableau suivant pour plus de détails sur les propriétés Spark associées à cette fonctionnalité.

Propriétés de la partition de coalescence adaptative Spark

Propriété	Valeur par défaut	Description
<code>spark.sql.adaptive.coalescePartitions.enabled</code>	<code>true</code> , sauf si <code>spark.sql.shuffle.partitions</code> est explicitement défini.	Lorsque <code>true</code> et <code>spark.sql.adaptive.enabled</code> est <code>true</code> , Spark fusionne les partitions réorganisées contiguës en fonction de la taille cible (spécifiée par <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>), afin d'éviter un trop grand nombre de petites tâches.
<code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>	64 Mo	Taille conseillée en octets de partition shuffle lors de la fusion. Cette configura

Propriété	Valeur par défaut	Description
		tion n'a d'effet que lorsque <code>spark.sql.adaptive.enabled</code> et <code>spark.sql.adaptive.coalescePartitions.enabled</code> sont tous deux <code>true</code> .
<code>spark.sql.adaptive.coalescePartitions.minPartitionNum</code>	25	Le nombre minimum de partitions shuffle après fusion. Cette configuration n'a d'effet que lorsque <code>spark.sql.adaptive.enabled</code> et <code>spark.sql.adaptive.coalescePartitions.enabled</code> sont tous deux <code>true</code> .
<code>spark.sql.adaptive.coalescePartitions.initialPartitionNum</code>	1 000	Nombre initial de partitions réorganisées avant la fusion. Cette configuration n'a d'effet que lorsque <code>spark.sql.adaptive.enabled</code> et <code>spark.sql.adaptive.coalescePartitions.enabled</code> sont tous deux <code>true</code> .

Nettoyage dynamique de partition

Le nettoyage dynamique de partition améliore les performances de tâche avec plus de précision en sélectionnant des partitions spécifiques au sein d'une table qui doivent être lues et traitées pour une requête spécifique. En réduisant la quantité de données lues et traitées, l'exécution du travail est nettement moins longue. Avec Amazon EMR 5.26.0, cette fonctionnalité est activée par défaut. Avec Amazon EMR 5.24.0 et 5.25.0, vous pouvez activer cette fonctionnalité en définissant la propriété

Spark depuis Spark ou lors `spark.sql.dynamicPartitionPruning.enabled` de la création de clusters.

Propriétés des partitions d'élimination dynamique Spark

Propriété	Valeur par défaut	Description
<code>spark.sql.dynamicPartitionPruning.enabled</code>	<code>true</code>	Lorsque c'est vrai, activez l'élimination dynamique des partitions.
<code>spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse</code>	<code>true</code>	Si la valeur est <code>true</code> , Spark effectue une vérification défensive avant l'exécution de la requête pour s'assurer que la réutilisation des échanges de diffusion dans les filtres d'élimination dynamique n'est pas interrompue par des règles de préparation ultérieures, telles que les règles en colonnes définies par l'utilisateur. Lorsque la réutilisation est interrompue et que cette configuration est <code>true</code> , Spark supprime les filtres d'élimination dynamique concernés afin de se prémunir contre les problèmes de performance et d'exactitude. Des problèmes d'exactitude peuvent survenir lorsque l'échange de diffusion du filtre d'élimination dynamique produit des résultats différents et incohérents par rapport à l'échange de diffusion de l'opération de jointure

Propriété	Valeur par défaut	Description
		correspondante. La définition de cette configuration sur <code>false</code> doit être effectuée avec prudence ; elle permet de contourner les scénarios, par exemple lorsque la réutilisation est interrompue par des règles en colonnes définies par l'utilisateur. Lorsque l'exécution adaptative des requêtes est activée, la réutilisation des diffusions est toujours appliquée.

Cette optimisation améliore les fonctionnalités existantes de Spark 2.4.2, qui prend uniquement en charge la transmission des prédicats statiques qui peuvent être résolus au moment voulu.

Voici des exemples de prédicat statiques en mode push Spark 2.4.2.

```
partition_col = 5
partition_col IN (1,3,5)
partition_col between 1 and 3
partition_col = 1 + 3
```

Le nettoyage dynamique de partition permet au moteur Spark de déduire dynamiquement au moment de l'exécution les partitions qui doivent être lues et qui peuvent être éliminées. Par exemple, la requête suivante implique deux tables : `store_sales` (qui contient toutes les ventes totales pour tous les magasins et est partitionnée par région) et `store_regions` (qui contient un mappage de régions pour chaque pays). Les tables contiennent des données sur les magasins répartis dans le monde entier, mais nous interrogeons uniquement des données pour l'Amérique du Nord.

```
select ss.quarter, ss.region, ss.store, ss.total_sales
from store_sales ss, store_regions sr
```

```
where ss.region = sr.region and sr.country = 'North America'
```

Sans nettoyage dynamique de partition, cette requête lit toutes les régions avant de filtrer le sous-ensemble de régions qui correspondent aux résultats de la sous-requête. Avec le nettoyage dynamique de partition, cette requête lit et traite uniquement les partitions pour les régions renvoyées dans la sous-requête. Cela permet d'économiser du temps et des ressources en lisant moins de données dans le stockage et en traitant moins d'enregistrements.

Aplatissement des sous-requêtes scalaires

Cette optimisation améliore les performances des requêtes qui ont des sous-requêtes scalaires sur la même table. Avec Amazon EMR 5.26.0, cette fonctionnalité est activée par défaut. Avec Amazon EMR 5.24.0 et 5.25.0, vous pouvez l'activer en définissant la propriété Spark depuis Spark ou lors `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` de la création de clusters. Lorsque cette propriété a la valeur `true`, l'optimiseur de requête aplatit les sous-requêtes scalaires regroupées qui utilisent la même relation si possible. Les sous-requêtes scalaires sont mises à plat en poussant tous les prédicats présents dans la sous-requête dans les fonctions d'agrégation, puis en effectuant une agrégation, avec toutes les fonctions d'agrégation, par relation.

Voici un exemple de requête qui bénéficie de cette optimisation.

```
select (select avg(age) from students          /* Subquery 1 */
        where age between 5 and 10) as group1,
       (select avg(age) from students          /* Subquery 2 */
        where age between 10 and 15) as group2,
       (select avg(age) from students          /* Subquery 3 */
        where age between 15 and 20) as group3
```

L'optimisation réécrit la requête précédente comme suit :

```
select c1 as group1, c2 as group2, c3 as group3
from (select avg (if(age between 5 and 10, age, null)) as c1,
           avg (if(age between 10 and 15, age, null)) as c2,
           avg (if(age between 15 and 20, age, null)) as c3 from students);
```

Notez que la requête réécrite lit la table des étudiants une seule fois et les prédicats des trois sous-requêtes sont ajoutés à la fonction `avg`.

DISTINCT avant INTERSECT

Cette optimisation optimise les jointures lors de l'utilisation INTERSECT. Avec Amazon EMR 5.26.0, cette fonctionnalité est activée par défaut. Avec Amazon EMR 5.24.0 et 5.25.0, vous pouvez l'activer en définissant la propriété Spark depuis Spark ou lors `spark.sql.optimizer.distinctBeforeIntersect.enabled` de la création de clusters. Les requêtes utilisant INTERSECT sont automatiquement converties pour utiliser une jointure semi-gauche. Lorsque cette propriété est définie sur `true`, l'optimiseur de requêtes redirige l'opérateur DISTINCT vers les enfants INTERSECT s'il détecte que l'opérateur peut associer le semi-gauche à un BroadcastHashJoin au lieu de `SortMergeJoin`.

Voici un exemple de requête qui bénéficie de cette optimisation.

```
(select item.brand brand from store_sales, item
  where store_sales.item_id = item.item_id)
intersect
(select item.brand cs_brand from catalog_sales, item
  where catalog_sales.item_id = item.item_id)
```

Sans activer cette propriété `spark.sql.optimizer.distinctBeforeIntersect.enabled`, la requête sera réécrite comme suit.

```
select distinct brand from
  (select item.brand brand from store_sales, item
    where store_sales.item_id = item.item_id)
left semi join
  (select item.brand cs_brand from catalog_sales, item
    where catalog_sales.item_id = item.item_id)
on brand <=> cs_brand
```

Lorsque vous activez cette propriété

`spark.sql.optimizer.distinctBeforeIntersect.enabled`, la requête est réécrite comme suit.

```
select brand from
  (select distinct item.brand brand from store_sales, item
    where store_sales.item_id = item.item_id)
left semi join
  (select distinct item.brand cs_brand from catalog_sales, item
    where catalog_sales.item_id = item.item_id)
```

```
on brand <=> cs_brand
```

Jonction de filtre Bloom

Cette optimisation peut améliorer les performances de certaines jonctions en préfiltrant un côté de la jonction à l'aide d'un [filtre Bloom](#) généré à partir des valeurs de l'autre côté de la jonction. Avec Amazon EMR 5.26.0, cette fonctionnalité est activée par défaut. Avec Amazon EMR 5.25.0, vous pouvez activer cette fonctionnalité en définissant la propriété Spark de manière `spark.sql.bloomFilterJoin.enabled` à ce qu'elle provienne `true` de Spark ou lors de la création de clusters.

L'exemple suivant est une requête qui peut bénéficier d'un filtre Bloom.

```
select count(*)
from sales, item
where sales.item_id = item.id
and item.category in (1, 10, 16)
```

Lorsque cette fonction est activée, le filtre Bloom est construit à partir de tous les ID d'éléments dont la catégorie figure dans l'ensemble des catégories interrogées. Pendant l'analyse de la table des ventes, le filtre Bloom est utilisé pour déterminer les ventes pour les éléments qui ne sont pas dans l'ensemble défini par le filtre Bloom. Ainsi, ces ventes identifiées peuvent être filtrées dès que possible.

Réorganisation optimisée des jonctions

Cette optimisation peut améliorer les performances des requêtes en réorganisant les jonctions impliquant des tables avec des filtres. Avec Amazon EMR 5.26.0, cette fonctionnalité est activée par défaut. Avec Amazon EMR 5.25.0, vous pouvez activer cette fonctionnalité en définissant le paramètre de configuration Spark sur `spark.sql.optimizer.sizeBasedJoinReorder.enabled true`. Par défaut, Spark procède à la jonction des tables de gauche à droite, comme indiqué dans la requête. Cette stratégie peut ignorer des possibilités d'exécution de jonctions plus petites avec des filtres dans un premier temps, afin de bénéficier des jonctions plus coûteuses ultérieurement.

L'exemple de requête ci-dessous indique tous les éléments renvoyés à partir de tous les magasins d'un pays. Sans réorganisation optimisée des jonctions, Spark joint les deux tables volumineuses `store_sales` et `store_returns` en premier, puis les joint avec `store` et finalement avec `item`.


```
select ss.item_value, sr.return_date, s.name, i.desc,
from store_sales ss, store_returns sr, store s, item i
where ss.id = sr.id and ss.store_id = s.id and ss.item_id = i.id
and s.country = 'USA'
```

Avec la réorganisation optimisée des jonctions, Spark joint d'abord `store_sales` et `store`, car `store` possède un filtre et est plus petite que `store_returns` et broadcastable. Ensuite, Spark joint `store_returns`, puis `item`. Si `item` disposait d'un filtre et pouvait être diffusée, elle pourrait bénéficier d'une réorganisation, ce qui entraînerait la jonction de `store_sales` avec `store`, puis avec `item`, et finalement avec `store_returns`.

Mise en cache des fragments de résultats Spark

Amazon EMR 6.6.0 et versions ultérieures incluent la fonctionnalité optionnelle Spark Result Fragment Caching qui met automatiquement en cache les fragments de résultats. Ces fragments de résultats font partie des résultats issus de sous-arborescences de requêtes stockées dans un compartiment Amazon S3 de votre choix. Les fragments de résultats de requête stockés sont réutilisés lors des exécutions de requêtes suivantes, ce qui permet d'accélérer les requêtes.

La mise en cache des fragments de résultats analyse vos SQL requêtes Spark et met en cache les fragments de résultats éligibles dans l'emplacement S3 que vous avez spécifié. Lors des exécutions de requête suivantes, les fragments de résultats de requête utilisables sont automatiquement détectés et extraits depuis S3. La mise en cache des fragments de résultats est différente de la mise en cache des ensembles de résultats, dans laquelle les requêtes suivantes doivent correspondre exactement à la requête d'origine pour renvoyer des résultats depuis le cache. Lorsqu'elle est utilisée pour des requêtes qui ciblent de manière répétée un sous-ensemble statique de vos données, la mise en cache des fragments de résultats accélère considérablement les performances.

Considérez la requête suivante, qui compte les commandes jusqu'en 2022 :

```
select
  l_returnflag,
  l_linestatus,
  count(*) as count_order
from
  lineitem
where
  l_shipdate <= current_date
```

```
and year(l_shipdate) == '2022'  
group by  
  l_returnflag,  
  l_linestatus
```

Au fil du temps, cette requête doit être exécutée tous les jours pour indiquer le total des ventes de l'année. Sans la mise en cache des fragments de résultats, les résultats pour tous les jours de l'année devront être recalculés chaque jour. La requête deviendra plus lente au fil du temps et sera plus lente à la fin de l'année, lorsque les 365 jours de résultats devront être recalculés.

Lorsque vous activez la mise en cache des fragments de résultats, vous utilisez les résultats de tous les jours précédents de l'année à partir du cache. Chaque jour, la fonctionnalité ne doit recalculer qu'un seul jour de résultats. Une fois que la fonctionnalité a calculé le fragment de résultat, elle met le fragment en cache. Par conséquent, les temps de requête activés par le cache sont rapides et restent constants pour chaque requête suivante.

Activation de la mise en cache des fragments de résultats Spark

Pour activer la mise en cache des fragments de résultats, veuillez suivre les étapes ci-dessous :

1. Créez un compartiment de cache dans Amazon S3 et autorisez l'accès en lecture/écriture pour EMRFS. Pour de plus amples informations, veuillez consulter [Autoriser l'accès aux EMRFS données dans Amazon S3](#).
2. Définissez la configuration Amazon EMR Spark pour activer la fonctionnalité.

```
spark.subResultCache.enabled = true  
spark.subResultCache.fs.root.path = s3://DOC-EXAMPLE-BUCKET/cache_dir/
```

3. Activez la gestion du cycle de vie S3 pour que le compartiment nettoie automatiquement les fichiers de cache.
4. Configurez éventuellement les propriétés `maxBufferSize` `reductionRatioThreshold` et pour affiner la fonctionnalité.

```
spark.sql.subResultCache.reductionRatioThreshold  
spark.sql.subResultCache.maxBufferSize
```

Considérations relatives à l'utilisation de la mise en cache des fragments de résultats

Les économies réalisées lorsque vous utilisez des résultats déjà mis en cache dans Amazon S3 plutôt que de les recalculer augmentent avec le nombre de fois où les mêmes résultats mis en cache peuvent être utilisés. Les requêtes comportant des analyses de grandes tables suivies de filtres ou d'agrégations de hachage qui réduisent la taille du résultat d'au moins 8 (c'est-à-dire un ratio d'au moins 8:1 dans la taille d'entrée:résultats) bénéficieront le plus de cette fonctionnalité. Plus le rapport de réduction entre l'entrée et les résultats est élevé, plus le rapport coût-bénéfice est important. Les requêtes présentant des taux de réduction inférieurs, mais comportant des étapes de calcul coûteuses entre l'analyse des tables et le filtrage ou les agrégations, en bénéficieront également, à condition que le coût de production des résultats soit supérieur à celui de leur extraction sur Amazon S3. Par défaut, la mise en cache des fragments de résultats ne prend effet que lorsqu'elle détecte qu'un taux de réduction sera d'au moins 8:1.

Lorsque vos requêtes réutilisent à plusieurs reprises les résultats mis en cache, les avantages de cette fonctionnalité sont les plus importants. Les requêtes de fenêtre progressives et incrémentielles en sont de bons exemples. Par exemple, une requête à fenêtre continue de 30 jours qui s'exécute déjà depuis 29 jours ne devrait extraire que 1/30e des données cibles de sa source d'entrée d'origine et utiliserait des fragments de résultats mis en cache pour les 29 jours précédents. Une requête de fenêtre incrémentielle en bénéficierait encore plus, car le début de la fenêtre reste fixe : à chaque invocation de la requête, un faible pourcentage du traitement nécessitera une lecture depuis la source d'entrée.

Voici d'autres considérations à prendre en compte lors de l'utilisation de la mise en cache des fragments de résultats :

- Les requêtes qui ne ciblent pas les mêmes données avec les mêmes fragments de requête auront un faible taux d'accès au cache et ne bénéficieront donc pas de cette fonctionnalité.
- Les requêtes présentant de faibles ratios de réduction qui ne contiennent pas d'étapes de calcul coûteuses produiront des résultats mis en cache à peu près aussi coûteux à lire qu'à traiter initialement.
- La première requête démontrera toujours une régression mineure en raison du coût d'écriture dans le cache.
- La fonctionnalité de mise en cache des fragments de résultats fonctionne exclusivement avec les fichiers Parquet. Les autres formats de fichiers ne sont pas pris en charge.

- Les tampons de la fonctionnalité de mise en cache des fragments de résultats essaieront uniquement de mettre en cache les scans dont la taille de fractionnement des fichiers est supérieure ou égale à 128 Mo. Avec la configuration par défaut de Spark, la mise en cache des fragments de résultats est désactivée si la taille de numérisation (taille totale de tous les fichiers analysés) divisée par le nombre de cœurs d'exécuteurs est inférieure à 128 Mo. Lorsque l'une des configurations Spark répertoriées ci-dessous est définie, la taille du fichier fractionné sera la suivante :

```
min(maxPartitionBytes, max(openCostInBytes, scan size / minPartitionNum))
```

- spark.sql. leafNodeDefaultParallélisme (la valeur par défaut est spark.default.parallelism)
- fichiers spark.sql. minPartitionNum (la valeur par défaut est spark.sql. leafNodeDefaultParallélisme)
- fichiers spark.sql. openCostInOctets
- fichiers spark.sql. maxPartitionBytes
- La fonction de mise en cache des fragments de résultats est mise en cache au niveau de la granularité de la RDD partition. Le ratio de réduction décrit précédemment, qui est par défaut de 8:1, est évalué par partition. RDD Les charges de travail dont les ratios par RDD réduction sont à la fois supérieurs et inférieurs à 8:1 peuvent présenter des avantages en termes de performances moindres que les charges de travail dont les ratios par RDD réduction sont constamment inférieurs à 8:1.
- La fonctionnalité de mise en cache des fragments de résultats utilise une mémoire tampon d'écriture de 16 Mo par défaut pour chaque RDD partition mise en cache. Si plus de 16 Mo sont mis en cache par RDD partition, le coût lié à la détermination de l'impossibilité d'écrire peut entraîner une régression des performances.
- Bien que, par défaut, Result Fragment Caching ne tente pas de mettre en cache les résultats des RDD partitions avec un ratio de réduction inférieur à 8:1 et plafonne sa mémoire tampon d'écriture à 16 Mo, ces deux valeurs sont réglables selon les configurations suivantes :

```
spark.sql.subResultCache.reductionRatioThreshold (default: 8.0)  
spark.sql.subResultCache.maxBufferSize (default: 16MB, max: 64MB)
```

- Plusieurs clusters utilisant la même EMR version d'Amazon peuvent partager le même emplacement de cache. Pour garantir l'exactitude des résultats, la mise en cache des fragments de résultats n'utilisera pas les résultats du cache écrits par différentes versions d'Amazon. EMR

- La mise en cache des fragments de résultats sera automatiquement désactivée pour les cas d'utilisation de Spark Streaming ou lorsque RecordServer Apache Ranger AWS Lake Formation est utilisé.
- Le fragment de résultat utilise le cache en lecture/écriture EMRFS et les compartiments Amazon S3. CSELe SSE KMS cryptage/SSES3/ est pris en charge.

Utiliser l'RAPIDSaccélérateur Nvidia pour Apache Spark

Avec les EMR versions 6.2.0 et ultérieures d'Amazon, vous pouvez utiliser le plugin [RAPIDSAccelerator for Apache Spark](#) de Nvidia pour accélérer Spark à l'aide de types d'instances d'unités de traitement EC2 graphique (GPU). RAPIDSAccelerator accélérera GPU vos pipelines de science des données Apache Spark 3.0 sans modifier le code et accélérera le traitement des données et la formation des modèles tout en réduisant considérablement les coûts d'infrastructure.

Les sections suivantes vous guident dans la configuration de votre EMR cluster pour utiliser le RAPIDS plug-in Spark pour Spark.

Choix des types d'instance

Pour utiliser le RAPIDS plugin Nvidia Spark- pour Spark, les groupes d'instances principaux et de tâches doivent utiliser des types d'EC2GPUinstances répondant aux [exigences matérielles](#) de Spark-RAPIDS. Pour consulter la liste complète des types d'GPUinstances EMR pris en charge par Amazon, consultez la section [Types d'instances pris en charge](#) dans le guide EMR de gestion Amazon. Le type d'instance du groupe d'instances principal peut être de type GPU ou non, mais GPU les types d'ARMinstance ne sont pas pris en charge.

Configuration des configurations d'applications pour votre cluster

1. Permettre EMR à Amazon d'installer les plug-ins sur votre nouveau cluster

Pour installer des plug-ins, fournissez la configuration suivante lors de la création de votre cluster :

```
{
  "Classification": "spark",
  "Properties": {
    "enableSparkRapids": "true"
  }
}
```

2. Configurer YARN pour utiliser GPU

Pour plus de détails sur l'utilisation de GPU onYARN, consultez la section [Utilisation de GPU on YARN dans la](#) documentation d'Apache Hadoop. Les exemples suivants présentent des exemples de YARN configurations pour les versions EMR 6.x et 7.x d'Amazon :

Amazon EMR 7.x

Exemple YARN de configuration pour Amazon EMR 7.x

```
{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.nodemanager.resource-plugins": "yarn.io/gpu",
    "yarn.resource-types": "yarn.io/gpu",
    "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices": "auto",
    "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables": "/usr/
bin",
    "yarn.nodemanager.linux-container-executor.cgroups.mount": "true",
    "yarn.nodemanager.linux-container-executor.cgroups.mount-path": "/spark-
rapids-cgroup",
    "yarn.nodemanager.linux-container-executor.cgroups.hierarchy": "yarn",
    "yarn.nodemanager.container-
executor.class": "org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
  }
}, {
  "Classification": "container-executor",
  "Properties": {

  },
  "Configurations": [
    {
      "Classification": "gpu",
      "Properties": {
        "module.enabled": "true"
      }
    },
    {
      "Classification": "cgroups",
      "Properties": {
        "root": "/spark-rapids-cgroup",
        "yarn-hierarchy": "yarn"
      }
    }
  ]
}
```

```

    }
  ]
}

```

Amazon EMR 6.x

Exemple YARN de configuration pour Amazon EMR 6.x

```

{
  "Classification": "yarn-site",
  "Properties": {
    "yarn.nodemanager.resource-plugins": "yarn.io/gpu",
    "yarn.resource-types": "yarn.io/gpu",
    "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices": "auto",
    "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables": "/usr/
bin",
    "yarn.nodemanager.linux-container-executor.cgroups.mount": "true",
    "yarn.nodemanager.linux-container-executor.cgroups.mount-path": "/sys/fs/
cgroup",
    "yarn.nodemanager.linux-container-executor.cgroups.hierarchy": "yarn",
    "yarn.nodemanager.container-
executor.class": "org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
  }
}, {
  "Classification": "container-executor",
  "Properties": {
  },
  "Configurations": [
    {
      "Classification": "gpu",
      "Properties": {
        "module.enabled": "true"
      }
    },
    {
      "Classification": "cgroups",
      "Properties": {
        "root": "/sys/fs/cgroup",
        "yarn-hierarchy": "yarn"
      }
    }
  ]
}

```

```
}
```

3. Configurer Spark pour utiliser RAPIDS

Voici les configurations requises pour permettre à Spark d'utiliser RAPIDS le plugin :

```
{
  "Classification": "spark-defaults",
  "Properties": {
    "spark.plugins": "com.nvidia.spark.SQLPlugin",
    "spark.executor.resource.gpu.discoveryScript": "/usr/lib/spark/scripts/gpu/getGpusResources.sh",
    "spark.executor.extraLibraryPath": "/usr/local/cuda/targets/x86_64-linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native"
  }
}
```

[XGBoost4J-La bibliothèque Spark](#) dans XGBoost la documentation est également disponible lorsque le RAPIDS plugin Spark est activé sur votre cluster. Vous pouvez utiliser la configuration suivante pour intégrer XGBoost votre tâche Spark :

```
{
  "Classification": "spark-defaults",
  "Properties": {
    "spark.submit.pyFiles": "/usr/lib/spark/jars/xgboost4j-spark_3.0-1.4.2-0.3.0.jar"
  }
}
```

Pour des configurations Spark supplémentaires que vous pouvez utiliser pour régler un EMR cluster GPU accéléré, reportez-vous au [guide de réglage de Rapids Accelerator pour Apache Spark](#) dans la documentation de Nvidia.github.io.

4. Configurer le YARN planificateur de capacité

DominantResourceCalculator doit être configuré pour permettre la GPU planification et l'isolation. Pour plus d'informations, reportez-vous à la section [Utiliser GPU sur YARN dans la documentation](#) d'Apache Hadoop.

```
{
```



```

"Classification":"capacity-scheduler",
"Properties":{
  "yarn.scheduler.capacity.resource-
calculator":"org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
}
}

```

5. Créez un JSON fichier pour inclure vos configurations

Vous pouvez créer un JSON fichier contenant votre configuration pour utiliser le RAPIDS plugin pour votre cluster Spark. Vous fournirez le fichier ultérieurement lors du lancement de votre cluster.

Le fichier peut être stocké localement ou sur S3. Pour plus d'informations sur la manière de fournir des configurations d'applications pour vos clusters, consultez [Configuration des applications](#).

Utilisez les exemples de fichiers suivants comme modèles pour créer vos propres configurations.

Amazon EMR 7.x

Exemple de **my-configurations.json** fichier pour Amazon EMR 7.x

```

[
  {
    "Classification":"spark",
    "Properties":{
      "enableSparkRapids":"true"
    }
  },
  {
    "Classification":"yarn-site",
    "Properties":{
      "yarn.nodemanager.resource-plugins":"yarn.io/gpu",
      "yarn.resource-types":"yarn.io/gpu",
      "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices":"auto",
      "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables":"/usr/bin",
      "yarn.nodemanager.linux-container-executor.cgroups.mount":"true",
      "yarn.nodemanager.linux-container-executor.cgroups.mount-path":"/spark-rapids-cgroup",
      "yarn.nodemanager.linux-container-executor.cgroups.hierarchy":"yarn",
      "yarn.nodemanager.container-executor.class":"org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
    }
  },
]

```

```

{
  "Classification":"container-executor",
  "Properties":{

},
  "Configurations":[
    {
      "Classification":"gpu",
      "Properties":{
        "module.enabled":"true"
      }
    },
    {
      "Classification":"cgroups",
      "Properties":{
        "root":"/spark-rapids-cgroup",
        "yarn-hierarchy":"yarn"
      }
    }
  ]
},
{
  "Classification":"spark-defaults",
  "Properties":{
    "spark.plugins":"com.nvidia.spark.SQLPlugin",
    "spark.executor.resource.gpu.discoveryScript":"/usr/lib/spark/scripts/
gpu/getGpusResources.sh",
    "spark.executor.extraLibraryPath":"/usr/local/cuda/targets/x86_64-
linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/
cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/
native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native",
    "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-
spark_3.0-1.4.2-0.3.0.jar",
    "spark.rapids.sql.concurrentGpuTasks":"1",
    "spark.executor.resource.gpu.amount":"1",
    "spark.executor.cores":"2",
    "spark.task.cpus":"1",
    "spark.task.resource.gpu.amount":"0.5",
    "spark.rapids.memory.pinnedPool.size":"0",
    "spark.executor.memoryOverhead":"2G",
    "spark.locality.wait":"0s",
    "spark.sql.shuffle.partitions":"200",
    "spark.sql.files.maxPartitionBytes":"512m"
  }
}

```

```

    },
    {
      "Classification": "capacity-scheduler",
      "Properties": {
        "yarn.scheduler.capacity.resource-
calculator": "org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
      }
    }
  ]

```

Amazon EMR 6.x

my-configurations.json Fichier d'exemple pour Amazon EMR 6.x

```

[
  {
    "Classification": "spark",
    "Properties": {
      "enableSparkRapids": "true"
    }
  },
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.nodemanager.resource-plugins": "yarn.io/gpu",
      "yarn.resource-types": "yarn.io/gpu",
      "yarn.nodemanager.resource-plugins.gpu.allowed-gpu-devices": "auto",
      "yarn.nodemanager.resource-plugins.gpu.path-to-discovery-executables": "/
usr/bin",
      "yarn.nodemanager.linux-container-executor.cgroups.mount": "true",
      "yarn.nodemanager.linux-container-executor.cgroups.mount-path": "/sys/fs/
cgroup",
      "yarn.nodemanager.linux-container-executor.cgroups.hierarchy": "yarn",
      "yarn.nodemanager.container-
executor.class": "org.apache.hadoop.yarn.server.nodemanager.LinuxContainerExecutor"
    }
  },
  {
    "Classification": "container-executor",
    "Properties": {
      },
      "Configurations": [
        {

```

```

        "Classification": "gpu",
        "Properties": {
            "module.enabled": "true"
        }
    },
    {
        "Classification": "cgroups",
        "Properties": {
            "root": "/sys/fs/cgroup",
            "yarn-hierarchy": "yarn"
        }
    }
]
},
{
    "Classification": "spark-defaults",
    "Properties": {
        "spark.plugins": "com.nvidia.spark.SQLPlugin",
        "spark.executor.resource.gpu.discoveryScript": "/usr/lib/spark/scripts/gpu/getGpusResources.sh",
        "spark.executor.extraLibraryPath": "/usr/local/cuda/targets/x86_64-linux/lib:/usr/local/cuda/extras/CUPTI/lib64:/usr/local/cuda/compat/lib:/usr/local/cuda/lib:/usr/local/cuda/lib64:/usr/lib/hadoop/lib/native:/usr/lib/hadoop-lzo/lib/native:/docker/usr/lib/hadoop/lib/native:/docker/usr/lib/hadoop-lzo/lib/native",
        "spark.submit.pyFiles": "/usr/lib/spark/jars/xgboost4j-spark_3.0-1.4.2-0.3.0.jar",
        "spark.rapids.sql.concurrentGpuTasks": "1",
        "spark.executor.resource.gpu.amount": "1",
        "spark.executor.cores": "2",
        "spark.task.cpus": "1",
        "spark.task.resource.gpu.amount": "0.5",
        "spark.rapids.memory.pinnedPool.size": "0",
        "spark.executor.memoryOverhead": "2G",
        "spark.locality.wait": "0s",
        "spark.sql.shuffle.partitions": "200",
        "spark.sql.files.maxPartitionBytes": "512m"
    }
},
{
    "Classification": "capacity-scheduler",
    "Properties": {
        "yarn.scheduler.capacity.resource-calculator": "org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
    }
}

```

```
}  
]
```

Ajoutez une action d'amorçage pour votre cluster

Pour plus d'informations sur la façon de fournir des scripts d'action Bootstrap lorsque vous créez votre cluster, consultez les principes de [base des actions Bootstrap](#) dans le guide de EMRgestion Amazon.

Les exemples de scripts suivants montrent comment créer un fichier d'action bootstrap pour Amazon EMR 6.x et 7.x :

Amazon EMR 7.x

Exemple de **my-bootstrap-action.sh** fichier pour Amazon EMR 7.x

YARN Pour gérer les GPU ressources avec les versions EMR 7.x d'Amazon, vous devez procéder au montage manuel CGroup v1 sur votre cluster. Vous pouvez le faire à l'aide d'un script d'action d'amorçage, comme indiqué dans cet exemple.

```
#!/bin/bash  
set -ex  
  
sudo mkdir -p /spark-rapids-cgroup/devices  
sudo mount -t cgroup -o devices cgroupv1-devices /spark-rapids-cgroup/devices  
sudo chmod a+rwx -R /spark-rapids-cgroup
```

Amazon EMR 6.x

my-bootstrap-action.sh Fichier d'exemple pour Amazon EMR 6.x

Pour les versions Amazon EMR 6.x, vous devez ouvrir CGroup des autorisations YARN sur votre cluster. Pour ce faire, vous pouvez utiliser un script d'action d'amorçage, comme illustré dans cet exemple.

```
#!/bin/bash  
set -ex  
  
sudo chmod a+rwx -R /sys/fs/cgroup/cpu,cpuacct
```

```
sudo chmod a+rwx -R /sys/fs/cgroup/devices
```

Lancement de votre cluster

La dernière étape consiste à lancer votre cluster avec les configurations de cluster mentionnées ci-dessus. Voici un exemple de commande pour lancer un cluster depuis Amazon EMR CLI :

```
aws emr create-cluster \  
--release-label emr-7.2.0 \  
--applications Name=Hadoop Name=Spark \  
--service-role EMR_DefaultRole_V2 \  
--ec2-attributes KeyName=my-key-pair,InstanceProfile=EMR_EC2_DefaultRole \  
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m4.4xlarge \  
                    InstanceGroupType=CORE,InstanceCount=1,InstanceType=g4dn.2xlarge \  
                    InstanceGroupType=TASK,InstanceCount=1,InstanceType=g4dn.2xlarge \  
--configurations file:///my-configurations.json \  
--bootstrap-actions Name='My Spark Rapids Bootstrap action',Path=s3://my-bucket/my-  
bootstrap-action.sh
```

Accès au shell de Spark

Le shell Spark est basé sur le Scala REPL (Read-Eval-Print-Loop). Il vous permet de créer des programmes Spark de façon interactive et de soumettre du travail à l'infrastructure. Vous pouvez accéder au shell Spark en vous connectant au nœud principal SSH et en l'appelant `spark-shell`. Pour plus d'informations sur la connexion au nœud principal, consultez la section [Se connecter au nœud principal SSH à l'aide](#) de l'Amazon EMR Management Guide. Les exemples suivants utilisent les journaux d'accès au HTTP serveur Apache stockés dans Amazon S3.

Note

Le compartiment utilisé dans ces exemples est disponible pour les clients ayant accès à la région USA Est (Virginie du Nord).

Par défaut, le shell Spark crée son propre [SparkContext](#) objet appelé `sc`. Vous pouvez utiliser ce contexte s'il est requis dans le REPL. `sqlContext` est également disponible en coque et il s'agit d'un [HiveContext](#).

Exemple Utilisation du shell de Spark pour compter les occurrences d'une chaîne dans un fichier stocké dans Amazon S3

Cet exemple utilise `sc` pour lire un fichier texte stocké dans Amazon S3.

```
scala> sc
res0: org.apache.spark.SparkContext = org.apache.spark.SparkContext@404721db

scala> val textFile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/
impressions/dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark crée `textFile` la [structure de données](#) associée. Ensuite, l'exemple compte le nombre de lignes figurant dans le fichier journal avec la chaîne « `cartoonnetwork.com` » :

```
scala> val linesWithCartoonNetwork = textFile.filter(line =>
  line.contains("cartoonnetwork.com")).count()
linesWithCartoonNetwork: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at
  filter at <console>:23
<snip>
<Spark program runs>
scala> linesWithCartoonNetwork
res2: Long = 9
```

Exemple Utilisation du shell de Spark basé sur Python pour compter les occurrences d'une chaîne dans un fichier stocké dans Amazon S3

Spark inclut également un shell basé sur Python, `pyspark`, que vous pouvez utiliser pour créer un prototype de programmes Spark écrits en Python. Tout comme avec `spark-shell`, `invoke pyspark` sur le nœud principal ; il possède également le même [SparkContext](#) objet.

```
>>> sc
<pyspark.context.SparkContext object at 0x7fe7e659fa50>
>>> textfile = sc.textFile("s3://elasticmapreduce/samples/hive-ads/tables/impressions/
dt=2009-04-13-08-05/ec2-0-51-75-39.amazon.com-2009-04-13-08-05.log")
```

Spark crée `textFile` la [structure de données](#) associée. Ensuite, l'exemple compte le nombre de lignes figurant dans le fichier journal avec la chaîne « `cartoonnetwork.com` ».

```
>>> linesWithCartoonNetwork = textfile.filter(lambda line: "cartoonnetwork.com" in
  line).count()
```

```
15/06/04 17:12:22 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library from the
embedded binaries
15/06/04 17:12:22 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo
library [hadoop-lzo rev EXAMPLE]
15/06/04 17:12:23 INFO fs.EmrFileSystem: Consistency disabled, using
com.amazon.ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation
<snip>
<Spark program continues>
>>> linesWithCartoonNetwork
9
```

Utiliser Amazon SageMaker Spark pour le machine learning

Lorsque vous utilisez Amazon EMR version 5.11.0 ou ultérieure, le `aws-sagemaker-spark-sdk` composant est installé avec Spark. Ce composant installe Amazon SageMaker Spark et les dépendances associées pour l'intégration de Spark à [Amazon SageMaker](#). Vous pouvez utiliser Amazon SageMaker Spark pour créer des pipelines d'apprentissage automatique (ML) Spark à l'aide d'Amazon SageMaker Stages. Pour plus d'informations, consultez le guide [Amazon SageMaker Spark README](#) sur GitHub et son [utilisation avec Apache Spark avec Amazon SageMaker](#) dans le manuel du SageMaker développeur Amazon.

Écriture d'une application Spark

Les applications [Spark](#) peuvent être écrites en Scala, Java ou Python. Il existe plusieurs exemples d'applications Spark proposés dans la rubrique [Exemples Spark](#) de la documentation Apache Spark. L'exemple d'estimation du nombre Pi est présenté ci-dessous dans les trois applications prises en charge en mode natif. Vous pouvez également consulter des exemples complets dans `$SPARK_HOME/examples` et sur [GitHub](#). Pour plus d'informations sur la façon de compiler JARs pour Spark, consultez la rubrique [Démarrage rapide](#) dans la documentation d'Apache Spark.

Scala

Pour éviter les problèmes de compatibilité avec Scala, nous vous conseillons d'utiliser les dépendances Spark pour obtenir la bonne version de Scala lorsque vous compilez une application Spark pour un cluster AmazonEMR. La version de Scala que vous devez utiliser dépend de la version de Spark installée sur votre cluster. Par exemple, la EMR version 5.30.1 d'Amazon utilise Spark 2.4.5, qui est construit avec Scala 2.11. Si votre cluster utilise la EMR version 5.30.1 d'Amazon, utilisez les dépendances Spark pour Scala 2.11. Pour plus d'informations sur les versions de Scala utilisées par Spark, consultez la [documentation d'Apache Spark](#).


```

package org.apache.spark.examples
import scala.math.random
import org.apache.spark._

/** Computes an approximation to pi */
object SparkPi {
  def main(args: Array[String]) {
    val conf = new SparkConf().setAppName("Spark Pi")
    val spark = new SparkContext(conf)
    val slices = if (args.length > 0) args(0).toInt else 2
    val n = math.min(100000L * slices, Int.MaxValue).toInt // avoid overflow
    val count = spark.parallelize(1 until n, slices).map { i =>
      val x = random * 2 - 1
      val y = random * 2 - 1
      if (x*x + y*y < 1) 1 else 0
    }.reduce(_ + _)
    println("Pi is roughly " + 4.0 * count / n)
    spark.stop()
  }
}

```

Java

```

package org.apache.spark.examples;

import org.apache.spark.SparkConf;
import org.apache.spark.api.java.JavaRDD;
import org.apache.spark.api.java.JavaSparkContext;
import org.apache.spark.api.java.function.Function;
import org.apache.spark.api.java.function.Function2;

import java.util.ArrayList;
import java.util.List;

/**
 * Computes an approximation to pi
 * Usage: JavaSparkPi [slices]
 */
public final class JavaSparkPi {

  public static void main(String[] args) throws Exception {
    SparkConf sparkConf = new SparkConf().setAppName("JavaSparkPi");

```

```
JavaSparkContext jsc = new JavaSparkContext(sparkConf);

int slices = (args.length == 1) ? Integer.parseInt(args[0]) : 2;
int n = 100000 * slices;
List<Integer> l = new ArrayList<Integer>(n);
for (int i = 0; i < n; i++) {
    l.add(i);
}

JavaRDD<Integer> dataSet = jsc.parallelize(l, slices);

int count = dataSet.map(new Function<Integer, Integer>() {
    @Override
    public Integer call(Integer integer) {
        double x = Math.random() * 2 - 1;
        double y = Math.random() * 2 - 1;
        return (x * x + y * y < 1) ? 1 : 0;
    }
}).reduce(new Function2<Integer, Integer, Integer>() {
    @Override
    public Integer call(Integer integer, Integer integer2) {
        return integer + integer2;
    }
});

System.out.println("Pi is roughly " + 4.0 * count / n);

jsc.stop();
}
```

Python

```
import argparse
import logging
from operator import add
from random import random

from pyspark.sql import SparkSession

logger = logging.getLogger(__name__)
logging.basicConfig(level=logging.INFO, format="%(levelname)s: %(message)s")
```

```
def calculate_pi(partitions, output_uri):
    """
    Calculates pi by testing a large number of random numbers against a unit circle
    inscribed inside a square. The trials are partitioned so they can be run in
    parallel on cluster instances.

    :param partitions: The number of partitions to use for the calculation.
    :param output_uri: The URI where the output is written, typically an Amazon S3
                       bucket, such as 's3://example-bucket/pi-calc'.
    """

    def calculate_hit(_):
        x = random() * 2 - 1
        y = random() * 2 - 1
        return 1 if x**2 + y**2 < 1 else 0

    tries = 100000 * partitions
    logger.info(
        "Calculating pi with a total of %s tries in %s partitions.", tries, partitions
    )
    with SparkSession.builder.appName("My PyPi").getOrCreate() as spark:
        hits = (
            spark.sparkContext.parallelize(range(tries), partitions)
            .map(calculate_hit)
            .reduce(add)
        )
        pi = 4.0 * hits / tries
        logger.info("%s tries and %s hits gives pi estimate of %s.", tries, hits, pi)
        if output_uri is not None:
            df = spark.createDataFrame([(tries, hits, pi)], ["tries", "hits", "pi"])
            df.write.mode("overwrite").json(output_uri)

if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument(
        "--partitions",
        default=2,
        type=int,
        help="The number of parallel partitions to use when calculating pi.",
    )
    parser.add_argument(
```

```
    "--output_uri", help="The URI where output is saved, typically an S3 bucket."
)
args = parser.parse_args()

calculate_pi(args.partitions, args.output_uri)
```

Améliorer les performances de Spark avec Amazon S3

Amazon EMR propose des fonctionnalités permettant d'optimiser les performances lors de l'utilisation de Spark pour interroger, lire et écrire des données enregistrées dans Amazon S3.

[S3 Select](#) peut améliorer les performances des requêtes CSV et JSON des fichiers dans certaines applications en « transférant » le traitement vers Amazon S3.

Le committer EMRFS optimisé pour S3 est une alternative à la [OutputCommitter](#) classe, qui utilise la fonctionnalité de téléchargement partitionné de pour améliorer les performances lors de l'écriture de fichiers EMRFS Parquet sur Amazon S3 à l'aide de SparkSQL, et Datasets. DataFrames

Rubriques

- [Utilisation de S3 Select avec Spark pour améliorer les performances des requêtes](#)
- [Utilisez le transmetteur EMRFS optimisé pour S3](#)
- [Utilisez le protocole de EMRFS validation optimisé pour S3](#)
- [Réessayez les requêtes Amazon S3 avec EMRFS](#)

Utilisation de S3 Select avec Spark pour améliorer les performances des requêtes

Avec les EMR versions 5.17.0 et ultérieures d'Amazon, vous pouvez utiliser [S3 Select](#) avec Spark sur Amazon. EMR S3 Select permet aux applications de ne récupérer qu'un sous-ensemble de données d'un objet. Pour AmazonEMR, le travail de calcul consistant à filtrer de grands ensembles de données à des fins de traitement est « transféré » du cluster vers Amazon S3, ce qui peut améliorer les performances de certaines applications et réduire la quantité de données transférées entre Amazon EMR et Amazon S3.

S3 Select est pris en charge avec CSV JSON les fichiers utilisant `s3selectCSV` et `s3selectJSON` les valeurs pour spécifier le format des données. Pour plus d'informations et d'exemples, consultez [Spécifier S3 Select dans votre code](#).

S3 Select est-il adapté à mon application ?

Nous vous recommandons de tester vos applications avec et sans S3 Select pour voir si son utilisation peut être adaptée à votre application.

Utilisez les consignes suivantes pour déterminer si votre application peut utiliser S3 Select :

- Votre requête exclut plus de la moitié de l'ensemble de données d'origine.
- Votre connexion réseau entre Amazon S3 et le EMR cluster Amazon offre une bonne vitesse de transfert et une bonne bande passante disponible. Amazon S3 ne compresse pas HTTP les réponses. La taille des réponses est donc susceptible d'augmenter pour les fichiers d'entrée compressés.

Considérations et restrictions

- Le chiffrement côté serveur Amazon S3 avec les clés de chiffrement fournies par le client (SSE-C) et le chiffrement côté client ne sont pas pris en charge.
- La propriété `AllowQuotedRecordDelimiters` n'est pas prise en charge. Si cette propriété est spécifiée, la requête échoue.
- Seuls CSV les JSON fichiers au format UTF -8 sont pris en charge. CSVsLes lignes multiples ne sont pas prises en charge.
- Seuls les fichiers non compressés ou gzip sont pris en charge.
- Spark CSV et JSON les options telles `quenanValue`, `positiveInfnegativeInf`, et les options liées aux enregistrements corrompus (par exemple, le mode `failfast` et `dropmalformed`) ne sont pas pris en charge.
- L'utilisation de virgules (,) avec des décimales n'est pas prise en charge. Par exemple, `10,000` n'est pas pris en charge, mais `10000` l'est.
- Les caractères de commentaire ne sont pas pris en charge dans la dernière ligne.
- Des lignes vides à la fin d'un fichier ne sont pas traitées.
- Les filtres suivants ne sont pas transmis à Amazon S3 :
 - Fonctions d'agrégation, telles que `COUNT()` et `SUM()`.

- Filtres qui CAST() un attribut. Par exemple, CAST(stringColumn as INT) = 1.
- Filtres avec un attribut d'objet ou complexe. Par exemple, intArray[1] = 1, objectColumn.objectNumber = 1.
- Filtres pour lesquels la valeur n'est pas une valeur littérale. Par exemple, intColumn1 = intColumn2
- Seuls [les types de données S3 Select pris en charge](#) sont pris en charge par les limitations de documentation.

Spécifier S3 Select dans votre code

Les exemples suivants montrent comment spécifier S3 Select pour CSV utiliser ScalaSQL, R et PySpark. Vous pouvez utiliser S3 Select JSON de la même manière. Pour une liste d'options, leurs valeurs par défaut et leurs limitations, consultez [Options](#).

PySpark

```
spark
  .read
  .format("s3selectCSV") // "s3selectJson" for Json
  .schema(...) // optional, but recommended
  .options(...) // optional
  .load("s3://path/to/my/datafiles")
```

R

```
read.df("s3://path/to/my/datafiles", "s3selectCSV", schema, header = "true",
        delimiter = "\t")
```

Scala

```
spark
  .read
  .format("s3selectCSV") // "s3selectJson" for Json
  .schema(...) // optional, but recommended
  .options(...) // optional. Examples:
  // .options(Map("quote" -> "\'", "header" -> "true")) or
  // .option("quote", "\'").option("header", "true")
  .load("s3://path/to/my/datafiles")
```

SQL

```
CREATE TEMPORARY VIEW MyView (number INT, name STRING) USING s3selectCSV OPTIONS
(path "s3://path/to/my/datafiles", header "true", delimiter "\t")
```

Options

Les options suivantes sont disponibles lors de l'utilisation de `s3selectCSV` et `s3selectJSON`. Si rien n'est spécifié, les valeurs par défaut sont utilisées.

Options avec S3select CSV

Option	Par défaut	Utilisation
<code>compression</code>	"none"	Indique si la compression est utilisée. "gzip" est le seul paramètre pris en charge, à part "none".
<code>delimiter</code>	","	Spécifie le caractère délimiteur de champ.
<code>quote</code>	'\"'	Spécifie le caractère de citation. La spécification d'une chaîne vide n'est pas prise en charge et entraîne une XML erreur mal formée.
<code>escape</code>	'\\'	Spécifie le caractère échappe.
<code>header</code>	"false"	"false" spécifie qu'il n'y a pas d'en-tête. "true" spécifie qu'une en-tête est dans la première ligne. Les en-têtes ne sont prises en charge que pour la première ligne, et les lignes vides avant une en-tête ne sont pas prises en charge.

Option	Par défaut	Utilisation
comment	"#"	Spécifie le caractère de commentaire. L'indicateur de commentaire ne peut pas être désactivé. En d'autres termes, la valeur de \u0000 n'est pas prise en charge.
nullValue	""	

Options avec S3select JSON

Option	Par défaut	Utilisation
compression	"none"	Indique si la compression est utilisée. "gzip" est le seul paramètre pris en charge, à part "none".
multiline	« false »	"false" indique que le format JSON est au LINES format S3 Select, ce qui signifie que chaque ligne des données d'entrée contient un seul JSON objet. "true" indique qu'il JSON est au DOCUMENT format S3 Select, ce qui signifie qu'un JSON objet peut s'étendre sur plusieurs lignes dans les données d'entrée.

Utilisez le transmetteur EMRFS optimisé pour S3

Le committer EMRFS optimisé pour S3 est une [OutputCommitter](#) implémentation alternative optimisée pour écrire des fichiers sur Amazon S3 lors de son utilisation. EMRFS Le validateur EMRFS

optimisé pour S3 améliore les performances des applications en évitant les opérations de liste et de renommage effectuées dans Amazon S3 pendant les phases de validation des tâches et des tâches. Le committer est disponible avec Amazon EMR version 5.19.0 et ultérieure, et est activé par défaut avec Amazon EMR 5.20.0 et versions ultérieures. Le committer est utilisé pour les tâches Spark qui utilisent Spark SQL ou Datasets. DataFrames À partir d'Amazon EMR 6.4.0, ce validateur peut être utilisé pour tous les formats courants ORC, y compris le parquet et les formats basés sur le texte (y compris et). CSV JSON Pour les versions antérieures à Amazon EMR 6.4.0, seul le format Parquet est pris en charge. Dans certaines situations, le valideur n'est pas utilisé. Pour de plus amples informations, veuillez consulter [Exigences relatives au transmetteur EMRFS optimisé pour S3](#).

Rubriques

- [Exigences relatives au transmetteur EMRFS optimisé pour S3](#)
- [Le validateur EMRFS optimisé pour S3 et les téléchargements partitionnés](#)
- [Considérations de réglage de tâche](#)
- [Activer le validateur EMRFS optimisé pour S3 pour Amazon 5.19.0 EMR](#)

Exigences relatives au transmetteur EMRFS optimisé pour S3

Le EMRFS validateur optimisé pour S3 est utilisé lorsque les conditions suivantes sont remplies :

- Vous exécutez des tâches Spark qui utilisent Spark SQL ou Datasets pour écrire des fichiers sur Amazon S3. DataFrames À partir d'Amazon EMR 6.4.0, ce validateur peut être utilisé pour tous les formats courants ORC, y compris le parquet et les formats basés sur le texte (y compris et). CSV JSON Pour les versions antérieures à Amazon EMR 6.4.0, seul le format Parquet est pris en charge.
- Les téléchargements partitionnés sont activés sur Amazon. EMR Il s'agit de l'option par défaut. Pour de plus amples informations, veuillez consulter [Le validateur EMRFS optimisé pour S3 et les téléchargements partitionnés](#).
- Le support de format de fichier intégré à Spark est utilisé. La prise en charge du format de fichier intégré est utilisée dans les circonstances suivantes :
 - Pour les tables Hive Metastore, le paramètre `spark.sql.hive.convertMetastoreParquet` est défini sur `true` pour les tables Parquet ou `spark.sql.hive.convertMetastoreOrc` sur les tables Orc `true` avec Amazon EMR 6.4.0 ou version ultérieure. Il s'agit des paramètres par défaut.

- Lorsque les tâches écrivent dans des sources de données ou des tables au format de fichier – par exemple, la table cible est créée avec la clause `USING parquet`.
- Lorsque des tâches écrivent dans des tableaux Parquet de metastore Hive non partitionnés. Le support Parquet intégré à Spark ne prend pas en charge avec les tableaux Hive partitionnés. Il s'agit d'une limitation connue. Pour plus d'informations, consultez la section [Conversion des tables Hive Metastore Parquet](#) dans le guide Apache Spark DataFrames et SQL Datasets.
- Les opérations de tâche Spark qui écrivent dans un emplacement de partition par défaut (par exemple, `/${table_location}/k1=v1/k2=v2/`) utilisent le valideur. Le valideur n'est pas utilisé si une opération de tâche écrit dans un emplacement de partition personnalisé, par exemple, si un emplacement de partition personnalisé est défini à l'aide de la commande `ALTER TABLE SQL`.
- Les valeurs suivantes pour Spark doivent être utilisées :
 - La propriété `spark.sql.parquet.fs.optimized.committer.optimization-enabled` doit être définie sur `true`. Il s'agit du paramètre par défaut d'Amazon EMR 5.20.0 et versions ultérieures. Avec Amazon EMR 5.19.0, la valeur par défaut est `false`. Pour plus d'informations sur la configuration de cette valeur, consultez [Activer le valideur EMRFS optimisé pour S3 pour Amazon 5.19.0 EMR](#).
 - Si vous écrivez dans des tables de métastore Hive non partitionnées, seuls les formats de fichier Parquet et Orc sont pris en charge. `spark.sql.hive.convertMetastoreParquet` doit être défini sur `true` si vous écrivez sur des tables de métastore Parquet Hive non partitionnées. `spark.sql.hive.convertMetastoreOrc` doit être défini sur `true` si vous écrivez dans des tables de métastore Orc Hive non partitionnées. Il s'agit des paramètres par défaut.
 - `spark.sql.parquet.output.committer.class` doit être défini sur `com.amazon.emr.committer.EmrOptimizedSparkSqlParquetOutputCommitter`. Il s'agit du paramètre par défaut.
 - `spark.sql.sources.commitProtocolClass` doit être réglé sur `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` ou `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` est le paramètre par défaut pour la série Amazon EMR 5.x version 5.30.0 et supérieure, et pour la série Amazon EMR 6.x version 6.2.0 et supérieure. `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` est le paramètre par défaut pour les EMR versions précédentes d'Amazon.
- Si des tâches Spark remplacent des ensembles de données Parquet partitionnés par des colonnes de partition dynamique, l'option d'écriture `partitionOverwriteMode` et

`spark.sql.sources.partitionOverwriteMode` doivent être définis sur `static`. Il s'agit du paramètre par défaut.

Note

L'option d'écriture `partitionOverwriteMode` a été introduite dans Spark 2.4.0. Pour la version 2.3.2 de Spark, incluse dans la EMR version 5.19.0 d'Amazon, définissez la propriété `spark.sql.sources.partitionOverwriteMode`

Occasions où le EMRFS validateur optimisé pour S3 n'est pas utilisé

En général, le EMRFS validateur optimisé pour S3 n'est pas utilisé dans les situations suivantes.

Situation	Pourquoi le validateur n'est pas utilisé
Lorsque vous écrivez à HDFS	Le committer prend uniquement en charge l'écriture sur Amazon S3 en utilisant EMRFS.
Lorsque vous utilisez le système de fichiers S3A	Le comité ne fait que soutenir EMRFS.
Lorsque vous utilisez MapReduce ou Spark RDD API	Le validateur prend uniquement en charge l'utilisation de Spark SQL ou APIs Dataset. DataFrame

Les exemples Scala suivants illustrent certaines situations supplémentaires qui empêchent le commit EMRFS optimisé pour S3 d'être utilisé en totalité (le premier exemple) et en partie (le second exemple).

Exemple – Mode de remplacement de partition dynamique

L'exemple Scala suivant indique à Spark d'utiliser un algorithme de validation différent, ce qui empêche complètement l'utilisation du validateur optimisé pour EMRFS S3. Le code définit la propriété `partitionOverwriteMode` sur `dynamic` pour n'écraser que les partitions sur lesquelles vous écrivez des données. Ensuite, les colonnes de partition dynamique sont spécifiées par `partitionBy`, et le mode d'écriture est défini sur `overwrite`.

```
val dataset = spark.range(0, 10)
```

```
.withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")
.option("partitionOverwriteMode", "dynamic")
.partitionBy("dt")
.parquet("s3://EXAMPLE-DOC-BUCKET/output")
```

Vous devez configurer les trois paramètres pour éviter d'utiliser le validateur EMRFS optimisé pour S3. Lorsque vous le faites, Spark exécute un algorithme de validation différent qui est spécifié dans le protocole de validation de Spark. Pour les versions d'Amazon EMR 5.x antérieures à la version 5.30.0 et pour les versions d'Amazon EMR 6.x antérieures à la version 6.2.0, le protocole de validation utilise le répertoire intermédiaire de Spark, qui est un répertoire temporaire créé sous l'emplacement de sortie commençant par `.spark-staging`. L'algorithme renomme séquentiellement les répertoires de partition, ce qui peut avoir un impact négatif sur les performances. Pour plus d'informations sur les EMR versions 5.30.0 et ultérieures d'Amazon et 6.2.0 et versions ultérieures, consultez [Utilisez le protocole de EMRFS validation optimisé pour S3](#)

L'algorithme dans Spark 2.4.0 exécute les étapes suivantes :

1. Les tentatives de tâches écrivent leur résultat dans les répertoires de partition situés dans le répertoire de préparation de Spark, par exemple, `${outputLocation}/spark-staging-${jobID}/k1=v1/k2=v2/`.
2. Pour chaque partition écrite, la tâche tente de suivre les chemins de partition relatifs, par exemple, `k1=v1/k2=v2`.
3. Lorsqu'une tâche se termine avec succès elle fournit le pilote avec tous les chemins de partition relatifs qu'elle a suivis.
4. Une fois toutes les tâches terminées, la phase de validation de tâche collecte tous les répertoires de partition que les tentatives de tâche réussies ont écrit dans le répertoire intermédiaire de Spark. Spark renomme de manière séquentielle chacun de ces répertoires dans son emplacement de sortie final à l'aide d'opérations attribution de nouveau nom de l'arborescence.
5. Le répertoire intermédiaire est supprimé avant que la phase de validation de tâche soit terminée.

Exemple – Emplacement de partition personnalisé

Dans cet exemple, le code Scala s'insère dans deux partitions. Une partition possède un emplacement de partition personnalisé. L'autre partition utilise l'emplacement de partition par défaut. Le validateur EMRFS optimisé pour S3 est uniquement utilisé pour écrire le résultat de la tâche sur la partition qui utilise l'emplacement de partition par défaut.

```

val table = "dataset"
val location = "s3://bucket/table"

spark.sql(s"""
  CREATE TABLE $table (id bigint, dt date)
  USING PARQUET PARTITIONED BY (dt)
  LOCATION '$location'
  """)

// Add a partition using a custom location
val customPartitionLocation = "s3://bucket/custom"
spark.sql(s"""
  ALTER TABLE $table ADD PARTITION (dt='2019-01-28')
  LOCATION '$customPartitionLocation'
  """)

// Add another partition using default location
spark.sql(s"ALTER TABLE $table ADD PARTITION (dt='2019-01-29')")

def asDate(text: String) = lit(text).cast("date")

spark.range(0, 10)
  .withColumn("dt",
    when($"id" > 4, asDate("2019-01-28")).otherwise(asDate("2019-01-29")))
  .write.insertInto(table)

```

Le code Scala crée les objets Amazon S3 suivants :

```

custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
custom_${folder$}
table/_SUCCESS
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
table/dt=2019-01-29_${folder$}
table_${folder$}

```

Lorsque vous écrivez des partitions dans des emplacements personnalisés, Spark utilise un algorithme de validation similaire à celui de l'exemple précédent, qui est décrit ci-dessous. Comme dans l'exemple précédent, l'algorithme se traduit par des attributions séquentielles de nouveaux noms, ce qui peut avoir un impact négatif sur les performances.

1. Lors de l'écriture de la sortie d'une partition dans un emplacement personnalisé, les tâches écrivent un fichier dans le répertoire intermédiaire de Spark, qui est créé sous l'emplacement de sortie final. Le nom du fichier inclut un caractère aléatoire UUID pour éviter les collisions de fichiers. La tentative de tâche suit chaque fichier, ainsi que le chemin de la sortie souhaité final.
2. Lorsqu'une tâche se termine avec succès, elle fournit au pilote les fichiers et les chemins de sortie souhaités finaux.
3. Une fois toutes les tâches terminées, la phase de validation de tâche renomme de manière séquentielle tous les fichiers qui ont été écrits pour les partitions dans les emplacements personnalisés en leurs chemins de sortie finaux.
4. Le répertoire intermédiaire est supprimé avant que la phase de validation de tâche soit terminée.

Le validateur EMRFS optimisé pour S3 et les téléchargements partitionnés

Pour utiliser le validateur EMRFS optimisé pour S3, vous devez activer les téléchargements partitionnés pour Amazon. EMR Les chargements partitionnés sont activés par défaut. Vous pouvez les réactiver si besoin est. Pour plus d'informations, consultez [Configurer le téléchargement partitionné pour Amazon S3](#) dans le guide de EMR gestion Amazon.

Le validateur EMRFS optimisé pour S3 utilise les caractéristiques transactionnelles des téléchargements partitionnés pour garantir que les fichiers écrits par des tentatives de tâche apparaissent uniquement dans l'emplacement de sortie de la tâche lors de la validation de la tâche. En utilisant les téléchargements partitionnés de cette manière, le validateur améliore les performances de validation des tâches par rapport à la version 2 de l' `FileOutputCommitter` algorithme par défaut. Lorsque vous utilisez le validateur EMRFS optimisé pour S3, il convient de prendre en compte certaines différences majeures par rapport au comportement de téléchargement partitionné traditionnel :

- Les chargements partitionnés sont toujours effectués, peu importe la taille du fichier. Cela diffère du comportement par défaut de EMRFS, où la `fs.s3n.multipart.uploads.split.size` propriété contrôle la taille du fichier à partir de laquelle les téléchargements partitionnés sont déclenchés.
- Les chargements partitionnés sont laissés en état non terminé pendant un plus long laps de temps jusqu'à ce que la tâche soit validée ou abandonnée. Cela diffère du comportement par défaut selon EMRFS lequel un téléchargement partitionné s'achève lorsqu'une tâche termine l'écriture d'un fichier donné.

En raison de ces différences, si un Spark Executor tombe en JVM panne ou est arrêté alors que des tâches sont en cours d'exécution et que des données sont écrites sur Amazon S3, les chargements partitionnés incomplets sont plus susceptibles d'être laissés pour compte. C'est pourquoi, lorsque vous utilisez le validateur EMRFS optimisé pour S3, veillez à suivre les meilleures pratiques en matière de gestion des téléchargements partitionnés ayant échoué. Pour plus d'informations, consultez la section [Meilleures pratiques relatives](#) à l'utilisation des compartiments Amazon S3 dans le Guide EMR de gestion Amazon.

Considérations de réglage de tâche

Le EMRFS validateur optimisé pour S3 consomme une petite quantité de mémoire pour chaque fichier écrit par une tentative de tâche jusqu'à ce que la tâche soit validée ou abandonnée. La quantité de mémoire consommée est négligeable dans la plupart des tâches. Pour les tâches dotées de tâches à longue exécution qui écrivent un grand nombre de fichiers, la mémoire consommée par le valideur peut être significative et nécessitera des ajustements à l'allocation de mémoire aux exécuteurs Spark. Vous pouvez ajuster la mémoire d'exécuteur à l'aide de la propriété `spark.executor.memory`. Une seule tâche écrivant 100 000 fichiers nécessiterait en général environ 100 Mo de mémoire supplémentaire. Pour plus d'informations, consultez [Propriétés d'applications](#) dans la Documentation de configuration Apache Spark.

Activer le validateur EMRFS optimisé pour S3 pour Amazon 5.19.0 EMR

Si vous utilisez Amazon EMR 5.19.0, vous pouvez définir manuellement la `spark.sql.parquet.fs.optimized.committer.optimization-enabled` propriété sur `true` lorsque vous créez un cluster ou depuis Spark si vous utilisez Amazon. EMR

Activation du commit EMRFS optimisé pour S3 lors de la création d'un cluster

Utilisez la classification de configuration `spark-defaults` pour définir la propriété `spark.sql.parquet.fs.optimized.committer.optimization-enabled` sur `true`. Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

Activation du EMRFS validateur optimisé pour S3 depuis Spark

Vous pouvez définir `spark.sql.parquet.fs.optimized.committer.optimization-enabled` sur `true` en l'ancrant dans un `SparkConf`, en le passant comme paramètre `--conf` dans le shell Spark ou les outils `spark-submit` et `spark-sql`, ou dans `conf/spark-defaults.conf`. Pour plus d'informations, consultez [Configuration Spark](#) dans la documentation Apache Spark.

L'exemple suivant montre comment activer le valideur en exécutant une commande `spark-sql`.

```
spark-sql \  
  --conf spark.sql.parquet.fs.optimized.committer.optimization-enabled=true \  
  -e "INSERT OVERWRITE TABLE target_table SELECT * FROM source_table;"
```

Utilisez le protocole de EMRFS validation optimisé pour S3

Le protocole de validation EMRFS optimisé pour S3 est une [FileCommitProtocol](#) implémentation alternative optimisée pour écrire des fichiers avec Spark Dynamic Partition Overwrite dans Amazon S3 lors de leur utilisation. EMRFS Le protocole améliore les performances des applications en évitant les opérations de renommage dans Amazon S3 pendant la phase de validation de la tâche de remplacement de partition dynamique Spark.

Notez que [Utilisez le transmetteur EMRFS optimisé pour S3](#) améliore également les performances en évitant les opérations de renommage. Cependant, cela ne fonctionne pas pour les cas de remplacement de partition dynamique, tandis que les améliorations du protocole de validation ne ciblent que les cas de remplacement de partition dynamique.

Le protocole de validation est disponible avec les EMR versions 5.30.0 et ultérieures d'Amazon et 6.2.0 et versions ultérieures et est activé par défaut. Amazon EMR a ajouté une amélioration du parallélisme à partir de la version 5.31.0. Le protocole est utilisé pour les tâches Spark qui utilisent Spark SQL ou Datasets. DataFrames Dans certaines circonstances, le protocole de validation n'est pas utilisé. Pour de plus amples informations, veuillez consulter [Exigences relatives au protocole de EMRFS validation optimisé pour S3](#).

Rubriques

- [Exigences relatives au protocole de EMRFS validation optimisé pour S3](#)
- [Le protocole de validation EMRFS optimisé pour S3 et les téléchargements partitionnés](#)
- [Considérations de réglage de tâche](#)

Exigences relatives au protocole de EMRFS validation optimisé pour S3

Le protocole de validation EMRFS optimisé pour S3 est utilisé lorsque les conditions suivantes sont remplies :

- Vous exécutez des tâches Spark qui utilisent Spark SQL ou Datasets pour remplacer les tables partitionnées. DataFrames
- Vous exécutez des tâches Spark dont le mode de remplacement de partition est `dynamic`.

- Les téléchargements partitionnés sont activés sur Amazon. EMR Il s'agit de l'option par défaut. Pour de plus amples informations, veuillez consulter [Le protocole de validation EMRFS optimisé pour S3 et les téléchargements partitionnés](#).
- Le cache du système de fichiers pour EMRFS est activé. Il s'agit de l'option par défaut. Vérifiez que le paramètre `fs.s3.impl.disable.cache` est défini sur `false`.
- Le support intégré des sources de données de Spark est utilisé. La prise en charge de la source de données intégrée est utilisée dans les circonstances suivantes :
 - Lorsque les tâches écrivent dans des sources de données ou des tables intégrées.
 - Lorsque les tâches écrivent dans la table Parquet du métastore Hive. Cela se produit lorsque `spark.sql.hive.convertInsertingPartitionedTable` et `spark.sql.hive.convertMetastoreParquet` sont tous deux définis sur `true`. Il s'agit des paramètres par défaut.
 - Lorsque les jobs écrivent dans la table de ORC métastore Hive. Cela se produit lorsque `spark.sql.hive.convertInsertingPartitionedTable` et `spark.sql.hive.convertMetastoreOrc` sont tous deux définis sur `true`. Il s'agit des paramètres par défaut.
- Les opérations de tâches Spark qui écrivent dans un emplacement de partition par défaut (par exemple, `${table_location}/k1=v1/k2=v2/`) utilisent le protocole de validation. Le protocole n'est pas utilisé si une opération de tâche écrit dans un emplacement de partition personnalisé (par exemple, si un emplacement de partition personnalisé est défini à l'aide de la commande `ALTER TABLE SQL`).
- Les valeurs suivantes pour Spark doivent être utilisées :
 - `spark.sql.sources.commitProtocolClass` doit être défini sur `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`. Il s'agit du paramètre par défaut pour les EMR versions 5.30.0 et ultérieures d'Amazon, ainsi que pour les versions 6.2.0 et supérieures.
 - L'option d'écriture `partitionOverwriteMode` ou `spark.sql.sources.partitionOverwriteMode` doit être définie sur `dynamic`. Le paramètre par défaut est `static`.

Note

L'option d'écriture `partitionOverwriteMode` a été introduite dans Spark 2.4.0. Pour la version 2.3.2 de Spark, incluse dans la EMR version 5.19.0 d'Amazon, définissez la propriété `spark.sql.sources.partitionOverwriteMode`

- Si les tâches Spark remplacent la table Parquet du métastore Hive, `spark.sql.hive.convertMetastoreParquet`, `spark.sql.hive.convertInsertingPartitionedTable` et `spark.sql.hive.convertMetastore.partitionOverwriteMode` doivent être définis sur `true`. Il s'agit des paramètres par défaut.
- Si les jobs Spark remplacent la ORC table de métastore Hive,, `spark.sql.hive.convertMetastoreOrc`, `spark.sql.hive.convertInsertingPartitionedTable` et `spark.sql.hive.convertMetastore.partitionOverwriteMode` doit être défini sur `true`. Il s'agit des paramètres par défaut.

Exemple – Mode de remplacement de partition dynamique

Dans cet exemple de Scala, l'optimisation est déclenchée. Tout d'abord, vous définissez la propriété `partitionOverwriteMode` sur `dynamic`. Cela remplace uniquement les partitions sur lesquelles vous écrivez des données. Ensuite, vous spécifiez les colonnes de partition dynamique avec `partitionBy` et définissez le mode d'écriture sur `overwrite`.

```
val dataset = spark.range(0, 10)
  .withColumn("dt", expr("date_sub(current_date(), id)"))

dataset.write.mode("overwrite")           // "overwrite" instead of "insert"
  .option("partitionOverwriteMode", "dynamic") // "dynamic" instead of "static"
  .partitionBy("dt")                       // partitioned data instead of
  unpartitioned data
  .parquet("s3://EXAMPLE-DOC-BUCKET/output") // "s3://" to use Amazon EMR file
  system, instead of "s3a://" or "hdfs://"
```

Lorsque le protocole de validation EMRFS optimisé pour S3 n'est pas utilisé

En général, le protocole de validation EMRFS optimisé pour S3 fonctionne de la même manière que le protocole de SQL validation Spark par défaut open source,.

`org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol`
L'optimisation ne se produira pas dans les situations suivantes.

Situation	Pourquoi le protocole de validation n'est pas utilisé
Lorsque vous écrivez à HDFS	Le protocole de validation prend uniquement en charge l'écriture sur Amazon S3 à l'aide de EMRFS.
Lorsque vous utilisez le système de fichiers S3A	Le protocole de validation ne prend en charge que EMRFS.
Lorsque vous utilisez MapReduce ou Spark RDD API	Le protocole de validation prend uniquement en charge l'utilisation de Spark SQL ou Dataset APIs. DataFrame
Lorsque le remplacement dynamique de la partition n'est pas déclenché	Le protocole de validation optimise uniquement les cas de remplacement dynamique de partitions. Pour les autres cas, consultez Utilisez le transmetteur EMRFS optimisé pour S3 .

Les exemples Scala suivants illustrent certaines situations supplémentaires auxquelles le protocole de validation EMRFS optimisé pour S3 délègue des tâches.
`SQLHadoopMapReduceCommitProtocol`

Exemple – Mode de remplacement dynamique de partition avec emplacement de partition personnalisé

Dans cet exemple, les programmes Scala remplacent deux partitions en mode de remplacement dynamique de partition. Une partition possède un emplacement de partition personnalisé. L'autre partition utilise l'emplacement de partition par défaut. Le protocole de validation EMRFS optimisé pour S3 améliore uniquement la partition qui utilise l'emplacement de partition par défaut.

```
val table = "dataset"
val inputView = "tempView"
val location = "s3://bucket/table"
```

```

spark.sql(s"""
  CREATE TABLE $table (id bigint, dt date)
  USING PARQUET PARTITIONED BY (dt)
  LOCATION '$location'
  """)

// Add a partition using a custom location
val customPartitionLocation = "s3://bucket/custom"
spark.sql(s"""
  ALTER TABLE $table ADD PARTITION (dt='2019-01-28')
  LOCATION '$customPartitionLocation'
  """)

// Add another partition using default location
spark.sql(s"ALTER TABLE $table ADD PARTITION (dt='2019-01-29')")

def asDate(text: String) = lit(text).cast("date")

spark.range(0, 10)
  .withColumn("dt",
    when($"id" > 4, asDate("2019-01-28")).otherwise(asDate("2019-01-29")))
  .createTempView(inputView)

// Set partition overwrite mode to 'dynamic'
spark.sql(s"SET spark.sql.sources.partitionOverwriteMode=dynamic")

spark.sql(s"INSERT OVERWRITE TABLE $table SELECT * FROM $inputView")

```

Le code Scala crée les objets Amazon S3 suivants :

```

custom/part-00001-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
custom_${folder$}
table/_SUCCESS
table/dt=2019-01-29/part-00000-035a2a9c-4a09-4917-8819-e77134342402.c000.snappy.parquet
table/dt=2019-01-29_${folder$}
table_${folder$}

```

Note

L'écriture sur des emplacements de partition personnalisés dans les versions antérieures de Spark peut entraîner une perte de données. Dans cet exemple, la partition `dt='2019-01-28'` serait perdue. Pour plus de détails, consultez [SPARK-35106](#). Ce

problème est résolu dans les EMR versions 5.33.0 et ultérieures d'Amazon, à l'exception des versions 6.0.x et 6.1.x.

Lorsque vous écrivez des partitions dans des emplacements personnalisés, Spark utilise un algorithme de validation similaire à celui de l'exemple précédent, qui est décrit ci-dessous. Comme dans l'exemple précédent, l'algorithme se traduit par des attributions séquentielles de nouveaux noms, ce qui peut avoir un impact négatif sur les performances.

L'algorithme dans Spark 2.4.0 exécute les étapes suivantes :

1. Lors de l'écriture de la sortie d'une partition dans un emplacement personnalisé, les tâches écrivent un fichier dans le répertoire intermédiaire de Spark, qui est créé sous l'emplacement de sortie final. Le nom du fichier inclut un caractère aléatoire UUID pour éviter les collisions de fichiers. La tentative de tâche suit chaque fichier, ainsi que le chemin de la sortie souhaité final.
2. Lorsqu'une tâche se termine avec succès, elle fournit au pilote les fichiers et les chemins de sortie souhaités finaux.
3. Une fois toutes les tâches terminées, la phase de validation de tâche renomme de manière séquentielle tous les fichiers qui ont été écrits pour les partitions dans les emplacements personnalisés en leurs chemins de sortie finaux.
4. Le répertoire intermédiaire est supprimé avant que la phase de validation de tâche soit terminée.

Le protocole de validation EMRFS optimisé pour S3 et les téléchargements partitionnés

Pour utiliser l'optimisation pour le remplacement dynamique des partitions dans le protocole de validation EMRFS optimisé pour S3, les téléchargements partitionnés doivent être activés sur Amazon. EMR Les chargements partitionnés sont activés par défaut. Vous pouvez les réactiver si besoin est. Pour plus d'informations, consultez [Configurer le téléchargement partitionné pour Amazon S3](#) dans le guide de EMR gestion Amazon.

Lors du remplacement dynamique de partitions, le protocole de validation EMRFS optimisé pour S3 utilise les caractéristiques transactionnelles des téléchargements partitionnés pour garantir que les fichiers écrits par des tentatives de tâche n'apparaissent dans l'emplacement de sortie de la tâche qu'au moment de la validation de la tâche. En utilisant les chargements partitionnés de cette manière, le protocole de validation améliore les performances de validation des tâches par rapport au protocole par défaut `SQLHadoopMapReduceCommitProtocol`. Lorsque vous utilisez le protocole

de validation EMRFS optimisé pour S3, il convient de prendre en compte certaines différences majeures par rapport au comportement de téléchargement partitionné traditionnel :

- Les chargements partitionnés sont toujours effectués, peu importe la taille du fichier. Cela diffère du comportement par défaut de EMRFS, où la `fs.s3n.multipart.uploads.split.size` propriété contrôle la taille du fichier à partir de laquelle les téléchargements partitionnés sont déclenchés.
- Les chargements partitionnés sont laissés en état non terminé pendant un plus long laps de temps jusqu'à ce que la tâche soit validée ou abandonnée. Cela diffère du comportement par défaut selon EMRFS lequel un téléchargement partitionné s'achève lorsqu'une tâche termine l'écriture d'un fichier donné.

En raison de ces différences, si un Spark Executor tombe en JVM panne ou est arrêté alors que des tâches sont en cours d'exécution et écrivent des données sur Amazon S3, ou si un pilote Spark JVM se bloque ou est arrêté pendant l'exécution d'une tâche, les téléchargements partitionnés incomplets sont plus susceptibles d'être laissés pour compte. C'est pourquoi, lorsque vous utilisez le protocole de validation EMRFS optimisé pour S3, veillez à suivre les meilleures pratiques en matière de gestion des téléchargements partitionnés ayant échoué. Pour plus d'informations, consultez la section [Meilleures pratiques relatives](#) à l'utilisation des compartiments Amazon S3 dans le Guide EMR de gestion Amazon.

Considérations de réglage de tâche

Sur les exécuteurs EMRFS Spark, le protocole de validation optimisé pour S3 consomme une petite quantité de mémoire pour chaque fichier écrit par une tentative de tâche jusqu'à ce que la tâche soit validée ou abandonnée. La quantité de mémoire consommée est négligeable dans la plupart des tâches.

Sur les pilotes EMRFS Spark, le protocole de validation optimisé pour S3 nécessite de la mémoire pour stocker les informations de métadonnées de chaque fichier validé jusqu'à ce que la tâche soit validée ou abandonnée. Dans la plupart des tâches, le paramètre de mémoire par défaut du pilote Spark est négligeable.

Pour les tâches qui ont des tâches de longue durée qui écrivent un grand nombre de fichiers, la mémoire que le protocole de validation consomme peut être perceptible et nécessiter des ajustements de la mémoire allouée pour Spark, en particulier pour les exécuteurs Spark. Vous pouvez régler la mémoire à l'aide de la propriété `spark.driver.memory` pour les pilotes Spark et de la propriété `spark.executor.memory` pour les exécuteurs Spark. Une seule tâche écrivant

100 000 fichiers nécessiterait en général environ 100 Mo de mémoire supplémentaire. Pour plus d'informations, consultez [Propriétés d'applications](#) dans la Documentation de configuration Apache Spark.

Réessayez les requêtes Amazon S3 avec EMRFS

Cette rubrique fournit des informations sur les stratégies de nouvelle tentative que vous pouvez utiliser lorsque vous envoyez des demandes à Amazon S3 avec EMRFS. Lorsque votre taux de demandes augmente, S3 essaie de s'adapter au nouveau taux. Au cours de ce processus, S3 peut limiter les demandes et renvoyer une erreur 503 Slow Down. Pour améliorer le taux de réussite de vos demandes S3, vous pouvez ajuster votre stratégie de réessai en configurant les propriétés dans votre configuration `emrfs-site`.

Vous pouvez ajuster votre stratégie de réessai de différentes manières.

- Augmentez la limite maximale de tentatives pour la stratégie de backoff exponentiel par défaut.
- Activez et configurez la stratégie de réessai additive-augmente/multiplicative-decrease (). AIMD AIMDest pris en charge pour les EMR versions 6.4.0 et ultérieures d'Amazon.

Utiliser la stratégie de backoff exponentiel par défaut

EMRFS utilise par défaut une stratégie d'interruption exponentielle pour réessayer les requêtes Amazon S3. La limite de EMRFS nouvelles tentatives par défaut est de 15. Pour éviter une erreur S3 503 Slow Down, vous pouvez augmenter le nombre limite de tentatives lorsque vous créez un nouveau cluster, sur un cluster en cours d'exécution ou lors de l'exécution de l'application.

Pour augmenter le nombre limite de tentatives, vous devez modifier la valeur de `fs.s3.maxRetries` dans votre configuration `emrfs-site`. L'exemple de configuration suivant définit `fs.s3.maxRetries` sur une valeur personnalisée de 30.

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.maxRetries": "30"
    }
  }
]
```

Pour plus d'informations sur l'utilisation des objets de configuration, consultez [Configuration des applications](#).

Utiliser la stratégie AIMD de nouvelle tentative

Avec les EMR versions 6.4.0 et ultérieures d'Amazon, EMRFS prend en charge une stratégie de nouvelle tentative alternative basée sur un modèle additive-augmentation/multiplicative-decrease (). AIMD La stratégie AIMD de nouvelle tentative est particulièrement utile lorsque vous travaillez avec de grands EMR clusters Amazon.

AIMD calcule un taux de demandes personnalisé à l'aide des données relatives aux récentes demandes réussies. Cette stratégie réduit le nombre de demandes limitées et le nombre total de tentatives requises par demande.

Pour activer la stratégie AIMD de nouvelle tentative, vous devez définir la `fs.s3.aimd.enabled` propriété sur `true` dans votre `emrfs-site` configuration, comme dans l'exemple suivant.

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.aimd.enabled": "true"
    }
  }
]
```

Pour plus d'informations sur l'utilisation des objets de configuration, consultez [Configuration des applications](#).

Paramètres de AIMD nouvelle tentative avancés

Vous pouvez configurer les propriétés répertoriées dans le tableau suivant pour affiner le comportement des nouvelles tentatives lorsque vous utilisez la stratégie de AIMD nouvelle tentative. Dans la plupart des cas d'utilisation, nous vous recommandons d'utiliser les valeurs par défaut.

Propriétés avancées AIMD de la stratégie de nouvelle tentative

Propriété	Valeur par défaut	Description
<code>fs.s3.aimd.increaseIncrement</code>	0.1	Contrôle la rapidité avec laquelle le taux de demandes

Propriété	Valeur par défaut	Description
		augmente lorsque des demandes consécutives aboutissent.
<code>fs.s3.aimd.reductionFactor</code>	2	Contrôle la rapidité avec laquelle le taux de demandes diminue lorsqu'Amazon S3 renvoie une réponse 503. Le facteur par défaut de 2 réduit le taux de demandes de moitié.
<code>fs.s3.aimd.minRate</code>	0.1	Définit la limite inférieure du taux de demandes lorsque les demandes sont soumises à une limitation soutenue par S3.
<code>fs.s3.aimd.initialRate</code>	5500	Définit le taux de demande initial, qui change ensuite en fonction des valeurs que vous spécifiez pour <code>fs.s3.aimd.increaseIncrement</code> et <code>fs.s3.aimd.reductionFactor</code> . Le taux initial est également utilisé pour les GET demandes, et ajusté proportionnellement (3500/5500) pour les demandes. PUT
<code>fs.s3.aimd.adjustWindow</code>	2	Contrôle la fréquence à laquelle le taux de demandes est ajusté, mesuré en nombre de réponses.

Propriété	Valeur par défaut	Description
<code>fs.s3.aimd.maxAttempts</code>	100	Définit le nombre maximum de tentatives pour essayer une demande.

Ajout d'une étape Spark

Vous pouvez utiliser EMR les étapes d'Amazon pour soumettre un travail au framework Spark installé sur un EMR cluster. Pour plus d'informations, consultez [les étapes](#) du guide EMR de gestion Amazon. Dans la console CLI, vous pouvez le faire à l'aide d'une étape de l'application Spark, qui exécute le `spark-submit` script en tant qu'étape en votre nom. Avec le API, vous utilisez une étape pour invoquer `spark-submit` en utilisant `command-runner.jar`.

Pour plus d'informations sur la soumission des applications à Spark, consultez la rubrique [Soumission d'applications](#) dans la documentation Apache Spark.

Pour soumettre une étape Spark à l'aide de la console


1. Ouvrez la EMR console Amazon à l'adresse <https://console.aws.amazon.com/emr>.
2. Dans la Liste de clusters, choisissez le nom de votre cluster.
3. Faites défiler l'affichage jusqu'à la section Étapes et développez-la, puis choisissez Ajouter une étape.
4. Dans la boîte de dialogue Ajouter une étape :
 - Pour Type d'étape, choisissez Spark application.
 - Pour Nom, acceptez le nom par défaut (application Spark) ou saisissez un nouveau nom.
 - Pour le Mode de déploiement, choisissez Client ou Cluster. Le mode client lance le programme pilote sur l'instance principale du cluster, tandis que le mode cluster lance votre programme pilote sur le cluster. En mode client, la sortie du journal du pilote apparaît dans les journaux des étapes, tandis qu'en mode cluster, la sortie du journal du pilote apparaît dans les journaux du premier YARN conteneur. Pour plus d'informations, consultez [Présentation du mode cluster](#) dans la documentation Apache Spark.
 - Spécifiez le paramètre Spark-submit options (Options Spark-envoi) souhaité. Pour plus d'informations sur les options `spark-submit`, consultez [Lancement d'applications avec spark-submit](#).

- Pour Emplacement de l'application, spécifiez le URI chemin local ou S3 de l'application.
 - Pour Arguments, laissez le champ vide.
 - Pour Action sur échec, acceptez l'option par défaut (Continuer).
5. Choisissez Ajouter. L'étape s'affiche dans la console avec le statut En suspens.
 6. Le statut de l'étape passe de Pending (En suspens) à Running (En cours d'exécution) puis à Completed (Terminé), au fur et à mesure de son exécution. Pour mettre à jour l'état, choisissez l'icône d'Actualisation au-dessus de la colonne Actions.
 7. Les résultats de l'étape se trouvent sur la page des détails du cluster de la EMR console Amazon à côté de votre étape, sous Fichiers journaux, si vous avez configuré la journalisation. Vous trouverez le cas échéant les informations sur l'étape dans le compartiment de journaux que vous avez configuré lorsque vous avez lancé le cluster.

Pour soumettre un travail à Spark à l'aide du AWS CLI

Soumettez une étape lorsque vous créez le cluster ou utilisez la sous-commande `aws emr add-steps` dans un cluster existant.

1. Utilisez `create-cluster` comme illustré dans l'exemple suivant.

 Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-7.2.0 --
applications Name=Spark \
--ec2-attributes KeyName=myKey --instance-type m5.xlarge --instance-count 3 \
--steps Type=Spark,Name="Spark Program",ActionOnFailure=CONTINUE,Args=[--
class,org.apache.spark.examples.SparkPi,/usr/lib/spark/examples/jars/spark-
examples.jar,10] --use-default-roles
```

Comme alternative, vous pouvez utiliser `command-runner.jar` comme illustré dans l'exemple suivant.

```
aws emr create-cluster --name "Add Spark Step Cluster" --release-label emr-7.2.0 \
```

```
--applications Name=Spark --ec2-attributes KeyName=myKey --instance-type m5.xlarge
--instance-count 3 \
--steps Type=CUSTOM_JAR,Name="Spark Program",Jar="command-
runner.jar",ActionOnFailure=CONTINUE,Args=[spark-example,SparkPi,10] --use-default-
roles
```

Note

Les caractères de continuation de ligne Linux (\) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

2. Vous pouvez également ajouter des étapes à un cluster déjà en cours d'exécution. Utilisez `add-steps`.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps
  Type=Spark,Name="Spark Program",ActionOnFailure=CONTINUE,Args=[--
class,org.apache.spark.examples.SparkPi,/usr/lib/spark/examples/jars/spark-
examples.jar,10]
```

Comme alternative, vous pouvez utiliser `command-runner.jar` comme illustré dans l'exemple suivant.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF --steps Type=CUSTOM_JAR,Name="Spark
  Program",Jar="command-runner.jar",ActionOnFailure=CONTINUE,Args=[spark-
example,SparkPi,10]
```

Pour soumettre un travail à Spark à l'aide du SDK pour Java

1. L'exemple suivant montre comment ajouter une étape à un cluster avec Spark à l'aide de Java.

```
AWSCredentials credentials = new BasicAWSCredentials(accessKey, secretKey);
AmazonElasticMapReduce emr = new AmazonElasticMapReduceClient(credentials);

StepFactory stepFactory = new StepFactory();
AmazonElasticMapReduceClient emr = new AmazonElasticMapReduceClient(credentials);
AddJobFlowStepsRequest req = new AddJobFlowStepsRequest();
req.withJobFlowId("j-1K48XXXXXXHCB");
```

```
List<StepConfig> stepConfigs = new ArrayList<StepConfig>();

HadoopJarStepConfig sparkStepConf = new HadoopJarStepConfig()
    .withJar("command-runner.jar")
    .withArgs("spark-submit", "--executor-memory", "1g", "--
class", "org.apache.spark.examples.SparkPi", "/usr/lib/spark/examples/jars/spark-
examples.jar", "10");

StepConfig sparkStep = new StepConfig()
    .withName("Spark Step")
    .withActionOnFailure("CONTINUE")
    .withHadoopJarStep(sparkStepConf);

stepConfigs.add(sparkStep);
req.withSteps(stepConfigs);
AddJobFlowStepsResult result = emr.addJobFlowSteps(req);
```

2. Affichez les résultats de l'étape en examinant les journaux de l'étape. Vous pouvez le faire AWS Management Console si vous avez activé la journalisation en choisissant Étapes, en sélectionnant votre étape, puis, pour les fichiers journaux, en choisissant l'une stdout ou l'autre des étapes stderr. Choisissez Afficher les journaux pour afficher les journaux disponibles.

Remplacement des paramètres de configuration par défaut de Spark

Vous souhaitez probablement remplacer les valeurs de configuration par défaut de Spark indépendamment pour chaque application. Vous pouvez faire cela lorsque vous soumettez des applications à l'aide d'une étape, ce qui transmet essentiellement des options à `spark-submit`. Par exemple, vous pouvez changer la mémoire allouée à un processus d'exécuteur en modifiant `spark.executor.memory`. Vous pouvez fournir le commutateur `--executor-memory` avec un argument tel que :

```
spark-submit --executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/
spark/examples/jars/spark-examples.jar 10
```

De même, vous pouvez ajuster `--executor-cores` et `--driver-memory`. Dans une étape, vous pouvez fournir les arguments suivants à l'étape :

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/
jars/spark-examples.jar 10
```

Vous pouvez également ajuster les paramètres qui n'ont peut-être pas de commutateur intégré à l'aide de l'option `--conf`. Pour plus d'informations sur les autres paramètres ajustables, consultez la rubrique [Chargement dynamique de propriétés Spark](#) dans la documentation Apache Spark.

Afficher l'historique de l'application Spark

Vous pouvez consulter les détails de Spark, de YARN l'application et de l'interface utilisateur de Tez à l'aide de l'onglet Interfaces utilisateur de l'application de la page détaillée d'un cluster dans la console. Les interfaces utilisateur (UI) des EMR applications Amazon vous permettent de résoudre et d'analyser plus facilement les tâches actives et l'historique des tâches.

Pour plus d'informations, consultez la section [Afficher l'historique des applications](#) dans le guide EMR de gestion Amazon.

Accédez au site Web Spark UIs

Vous pouvez consulter le site Web Spark UIs en suivant les procédures de création d'un SSH tunnel ou de création d'un proxy décrites dans la section intitulée [Connect to the cluster](#) du Amazon EMR Management Guide, puis en accédant à celle correspondant YARN ResourceManager à votre cluster. Choisissez le lien sous Tracking UI (UI de suivi) pour votre application. Si votre application est en cours d'exécution, vous voyez ApplicationMaster. Vous êtes dirigé vers l'interface utilisateur web du principal de l'application Spark sur le port 20888, quel que soit l'endroit où se trouve le pilote. Le pilote peut se trouver sur le nœud principal du cluster si vous l'exécutez en mode YARN client. Si vous exécutez une application en mode YARN cluster, le pilote se trouve dans le dossier correspondant ApplicationMaster à l'application sur le cluster. Si votre application est terminée, l'historique s'affiche, qui vous permet d'accéder au numéro de port Spark HistoryServer UI 18080 du nœud principal du EMR cluster. Ceci s'applique aux applications déjà terminées. Vous pouvez également accéder à l' HistoryServer interface utilisateur de Spark directement sur `http://master-public-dns-name:18080/`.

Avec les EMR versions 5.25.0 et ultérieures d'Amazon, vous pouvez accéder à l'interface utilisateur du serveur d'historique Spark depuis la console sans configurer de proxy Web via une SSH connexion. Pour plus d'informations, consultez [Afficher les interfaces utilisateur d'application persistante](#).

Utilisation du connecteur Amazon Kinesis Data Streams pour le streaming structuré Spark

Les EMR versions 7.1.0 et supérieures d'Amazon incluent un connecteur Amazon Kinesis Data Streams de streaming structuré Spark dans l'image de publication. Avec ce connecteur, vous pouvez utiliser Spark on Amazon EMR pour traiter les données stockées dans Amazon Kinesis Data Streams. Le connecteur prend en charge à la fois les types de consommation GetRecords (débit partagé) et SubscribeToShard (ventilation améliorée). Cette intégration est basée sur le [spark-sql-kinesis-connector](#). Pour plus de détails sur la façon de commencer à utiliser le connecteur, consultez le [README](#).

L'exemple suivant montre comment utiliser le connecteur pour lancer une application Spark avec Amazon EMR

```
spark-submit my_kinesis_streaming_script.py
```

Utilisation de l'intégration d'Amazon Redshift pour Apache Spark avec Amazon EMR

Dans les EMR versions 6.4.0 et ultérieures d'Amazon, chaque image de version inclut un connecteur entre [Apache Spark](#) et Amazon Redshift. Avec ce connecteur, vous pouvez utiliser Spark sur Amazon EMR pour traiter les données stockées dans Amazon Redshift. Pour les EMR versions 6.4.0 à 6.8.0 d'Amazon, l'intégration est basée sur le connecteur [spark-redshiftopen source](#). Pour les EMR versions 6.9.0 et ultérieures d'Amazon, l'[intégration Amazon Redshift pour Apache Spark](#) a été migrée de la version communautaire vers une intégration native.

Rubriques

- [Lancement d'une application Spark à l'aide de l'intégration Amazon Redshift pour Apache Spark](#)
- [Authentification avec l'intégration Amazon Redshift pour Apache Spark](#)
- [Lecture et écriture à partir de et vers Amazon Redshift](#)
- [Considérations et limites relatives à l'utilisation du connecteur Spark](#)

Lancement d'une application Spark à l'aide de l'intégration Amazon Redshift pour Apache Spark

Pour les EMR versions 6.4 à 6.9 d'Amazon, vous devez utiliser l'option `--packages` ou `--jars` pour spécifier lequel des JAR fichiers suivants vous souhaitez utiliser. L'option `--jars` spécifie les dépendances stockées localement sur HDFS, dans ou à l'aide de HTTP /S. Pour connaître les autres emplacements de fichiers pris en charge par l'option `--jars`, consultez la section [Gestion avancée des dépendances](#) dans la documentation de Spark. L'option `--packages` spécifie les dépendances stockées dans le référentiel public de Maven.

- `spark-redshift.jar`
- `spark-avro.jar`
- `RedshiftJDBC.jar`
- `minimal-json.jar`

Les versions 6.10.0 et ultérieures d'Amazon ne nécessitent pas cette `minimal-json.jar` dépendance et installent automatiquement les autres dépendances sur chaque cluster par défaut. Les exemples suivants montrent comment lancer une application Spark avec l'intégration Amazon Redshift pour Apache Spark.

Amazon EMR 6.10.0 +

L'exemple suivant montre comment lancer une application Spark avec le `spark-redshift` connecteur avec Amazon EMR versions 6.10 et supérieures.

```
spark-submit my_script.py
```

Amazon EMR 6.4.0 - 6.9.x

Pour lancer une application Spark avec le `spark-redshift` connecteur sur les EMR versions 6.4 à 6.9 d'Amazon, vous devez utiliser l'option `--packages` ou `--jars`, comme le montre l'exemple suivant. Notez que les chemins répertoriés avec l'option `--jars` sont les chemins par défaut pour les JAR fichiers.

```
spark-submit \  
  --jars /usr/share/aws/redshift/jdbc/RedshiftJDBC.jar,/usr/share/aws/redshift/  
  spark-redshift/lib/spark-redshift.jar,/usr/share/aws/redshift/spark-redshift/lib/  
  spark-avro.jar,/usr/share/aws/redshift/spark-redshift/lib/minimal-json.jar \  
  \
```



```
my_script.py
```

Authentification avec l'intégration Amazon Redshift pour Apache Spark

Utilisation AWS Secrets Manager pour récupérer les informations d'identification et se connecter à Amazon Redshift

L'exemple de code suivant montre comment récupérer les informations d'identification AWS Secrets Manager pour vous connecter à un cluster Amazon Redshift avec l' PySpark interface d'Apache Spark en Python.

```
from pyspark.sql import SQLContext
import boto3

sc = # existing SparkContext
sql_context = SQLContext(sc)

secretsmanager_client = boto3.client('secretsmanager')
secret_manager_response = secretsmanager_client.get_secret_value(
    SecretId='string',
    VersionId='string',
    VersionStage='string'
)
username = # get username from secret_manager_response
password = # get password from secret_manager_response
url = "jdbc:redshift://redshifthost:5439/database?user=" + username + "&password=" +
password

# Read data from a table
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "my_table") \
    .option("tempdir", "s3://path/for/temp/data") \
    .load()
```

Utilisation IAM pour récupérer les informations d'identification et se connecter à Amazon Redshift

Vous pouvez utiliser le pilote JDBC version 2 fourni par Amazon Redshift pour vous connecter à Amazon Redshift avec le connecteur Spark. Pour utiliser AWS Identity and Access Management (IAM), [configurez votre JDBC URL IAM authentication](#). Pour vous connecter à un cluster Redshift depuis AmazonEMR, vous devez autoriser votre IAM rôle à récupérer des informations d'identification temporaires IAM. Attribuez les autorisations suivantes à votre IAM rôle afin qu'il puisse récupérer les informations d'identification et exécuter les opérations Amazon S3.

- [Redshift : GetClusterCredentials](#) (pour les clusters Amazon Redshift provisionnés)
- [Redshift : DescribeClusters](#) (pour les clusters Amazon Redshift provisionnés)
- [Redshift : GetWorkgroup](#) (pour les groupes de travail Amazon Redshift Serverless)
- [Redshift : GetCredentials](#) (pour les groupes de travail Amazon Redshift Serverless)
- [s3 : GetBucket](#)
- [s3 : GetBucketLocation](#)
- [s3 : GetObject](#)
- [s3 : PutObject](#)
- [s3 : GetBucketLifecycleConfiguration](#)

Pour plus d'informations sur `GetClusterCredentials`, consultez [Stratégies de ressources pour GetClusterCredentials](#).

Vous devez également vous assurer qu'Amazon Redshift peut assumer ce IAM rôle pendant COPY et UNLOAD pendant les opérations.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "redshift.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

```
}
```

L'exemple suivant utilise IAM l'authentification entre Spark et Amazon Redshift :

```
from pyspark.sql import SQLContext
import boto3

sc = # existing SparkContext
sql_context = SQLContext(sc)

url = "jdbc:redshift:iam://redshift-host:redshift-port/db-name"
iam_role_arn = "arn:aws:iam::account-id:role/role-name"

# Read data from a table
df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("aws_iam_role", iam_role_arn) \
    .option("dbtable", "my_table") \
    .option("tempdir", "s3a://path/for/temp/data") \
    .mode("error") \
    .load()
```

Lecture et écriture à partir de et vers Amazon Redshift

Les exemples de code suivants permettent de lire et PySpark d'écrire des exemples de données depuis et vers une base de données Amazon Redshift avec source de données API et à l'aide de Spark. SQL

Data source API

PySpark À utiliser pour lire et écrire des exemples de données depuis et vers une base de données Amazon Redshift avec source de données. API

```
import boto3
from pyspark.sql import SQLContext

sc = # existing SparkContext
sql_context = SQLContext(sc)

url = "jdbc:redshift:iam://redshifthost:5439/database"
```

```
aws_iam_role_arn = "arn:aws:iam::accountID:role/roleName"

df = sql_context.read \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .load()

df.write \
    .format("io.github.spark_redshift_community.spark.redshift") \
    .option("url", url) \
    .option("dbtable", "tableName_copy") \
    .option("tempdir", "s3://path/for/temp/data") \
    .option("aws_iam_role", "aws_iam_role_arn") \
    .mode("error") \
    .save()
```

SparkSQL

PySpark À utiliser pour lire et écrire des exemples de données depuis et vers une base de données Amazon Redshift avec Spark. SQL

```
import boto3
import json
import sys
import os
from pyspark.sql import SparkSession

spark = SparkSession \
    .builder \
    .enableHiveSupport() \
    .getOrCreate()

url = "jdbc:redshift:iam://redshifthost:5439/database"
aws_iam_role_arn = "arn:aws:iam::accountID:role/roleName"

bucket = "s3://path/for/temp/data"
tableName = "tableName" # Redshift table name

s = f"""CREATE TABLE IF NOT EXISTS {tableName} (country string, data string)
USING io.github.spark_redshift_community.spark.redshift
```

```
OPTIONS (dbtable '{tableName}', tempdir '{bucket}', url '{url}', aws_iam_role
'{aws_iam_role_arn}' ); ""

spark.sql(s)

columns = ["country" ,"data"]
data = [("test-country", "test-data")]
df = spark.sparkContext.parallelize(data).toDF(columns)

# Insert data into table
df.write.insertInto(tableName, overwrite=False)
df = spark.sql(f"SELECT * FROM {tableName}")
df.show()
```

Considérations et limites relatives à l'utilisation du connecteur Spark

- Nous vous recommandons d'activer SSL la JDBC connexion entre Spark on Amazon EMR et Amazon Redshift.
- À titre de bonne pratique, nous vous recommandons de gérer les informations d'identification du cluster Amazon Redshift dans AWS Secrets Manager . Voir [Utiliser AWS Secrets Manager pour récupérer les informations d'identification pour se connecter à Amazon Redshift](#) pour un exemple.
- Nous vous recommandons de transmettre un IAM rôle avec le paramètre `aws_iam_role` d'authentification Amazon Redshift.
- Il `tempdir` URI pointe vers un emplacement Amazon S3. Ce répertoire temporaire n'est pas nettoyé automatiquement et peut donc entraîner des coûts supplémentaires.
- Tenez compte des recommandations suivantes pour Amazon Redshift :
 - Nous vous recommandons de bloquer l'accès public au cluster Amazon Redshift.
 - Nous vous recommandons d'activer la [journalisation des audits d'Amazon Redshift](#).
 - Nous vous recommandons d'activer le [chiffrement au repos d'Amazon Redshift](#).
- Tenez compte des recommandations suivantes pour Amazon S3 :
 - Nous vous recommandons de [bloquer l'accès public aux compartiments Amazon S3](#).
 - Nous vous recommandons d'utiliser le [chiffrement côté serveur Amazon S3](#) pour chiffrer les compartiments Amazon S3 utilisés.
 - Nous vous recommandons d'utiliser les [politiques de cycle de vie d'Amazon S3](#) pour définir les règles de conservation du compartiment Amazon S3.

- Amazon vérifie EMR toujours le code importé depuis l'open source dans l'image. Pour des raisons de sécurité, nous ne prenons pas en charge les méthodes d'authentification suivantes de Spark vers Amazon S3 :
 - Configuration des clés AWS d'accès dans la classification `hadoop-env` de configuration
 - Encodage des clés AWS d'accès dans le `tempdir` URI

Pour plus d'informations sur l'utilisation du connecteur et les paramètres qu'il prend en charge, consultez les ressources suivantes :

- [Intégration d'Amazon Redshift pour Apache Spark](#) dans le Guide de gestion Amazon Redshift
- Le [référentiel communautaire spark-redshift](#) sur Github

Historique des versions de Spark

Le tableau suivant répertorie la version de Spark incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Important

[Apache Spark version 2.3.1, disponible à partir de la EMR version 5.16.0 d'Amazon, adresses CVE-2018-8024 et -2018-1334. CVE](#) Nous vous recommandons de migrer les versions antérieures de Spark vers la version 2.3.1 ou ultérieure.

Informations sur la version de Spark

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-7.2.0	3.5.1	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode,

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
		hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.36.2	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-7.1.0	3.5.0	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-7.0.0	3.5.0	delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.15.0	3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.14.0	3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.13.0	3.4.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.12.0	3.4.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.11.1	3.3.2	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.11.0	3.3.2	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.10.1	3.3.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.10.0	3.3.1	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.9.1	3.3.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.9.0	3.3.0	aws-sagemaker-spark-sdk, delta, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.8.1	3.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.8.0	3.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.7.0	3.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.36.1	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.36.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.6.0	3.2.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.35.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.5.0	3.1.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.4.0	3.1.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.3.1	3.1.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.3.0	3.1.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.2.1	3.0.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.2.0	3.0.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.1.1	3.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.1.0	3.0.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-6.0.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-6.0.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.34.0	2.4.8	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.33.1	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.33.0	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.32.1	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.32.0	2.4.7	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.31.1	2.4.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.31.0	2.4.6	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.30.2	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.30.1	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.30.0	2.4.5	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-notebook-env, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
mr-5.29.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.28.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.28.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.27.1	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.27.0	2.4.4	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.26.0	2.4.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.25.0	2.4.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.24.1	2.4.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.24.0	2.4.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.23.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.23.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.22.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.21.2	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.21.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.21.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.20.1	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.20.0	2.4.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.19.1	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.19.0	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.18.1	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.18.0	2.3.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.17.2	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.17.1	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.17.0	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, emr-s3-select, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.16.1	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.16.0	2.3.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.15.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.15.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.14.2	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.14.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.14.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.13.1	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.13.0	2.3.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.12.3	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.12.2	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.12.1	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.12.0	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.11.4	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.11.3	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.11.2	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.11.1	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.11.0	2.2.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.10.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.10.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.9.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.9.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.8.3	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.8.2	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.8.1	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.8.0	2.2.0	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.7.1	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.7.0	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.6.1	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.6.0	2.1.1	emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.4	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.3	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.5.2	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.5.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.4.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.4.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.3.2	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.3.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.3.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.3	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.2.2	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.1	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.2.0	2.0.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.1.1	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.1.0	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.0.3	2.0.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-5.0.2	2.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.0.1	2.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-5.0.0	2.0.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-4.9.6	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.5	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.4	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-4.9.3	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.2	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.9.1	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-4.8.5	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.4	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.3	1.6.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-4.8.2	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.1	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.8.0	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-4.7.4	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.3	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.2	1.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-4.7.1	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.7.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.6.1	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-4.6.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.5.0	1.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.4.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-4.3.0	1.6.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.2.0	1.5.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave
emr-4.1.0	1.5.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-ht tpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resource manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Étiquette Amazon EMR Release	Version de Spark	Composants installés avec Spark
emr-4.0.0	1.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datano de, hadoop-hdfs-namenode, hadoop-ftpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave

Apache Sqoop

Apache Sqoop est un outil de transfert de données entre Amazon S3, Hadoop et des bases de données. HDFS RDBMS Pour plus d'informations, consultez le [site web Apache Sqoop](#). Sqoop est inclus dans les EMR versions 5.0.0 et ultérieures d'Amazon. Les versions antérieures incluent Sqoop en tant qu'application sandbox. Pour plus d'informations, consultez [Versions publiées par Amazon EMR 4.x](#).

Rubriques

- [Informations sur la version de Sqoop](#)
- [Considérations relatives à Sqoop sur Amazon EMR](#)
- [Historique des versions de Sqoop](#)

Informations sur la version de Sqoop

Version Sqoop pour 7.2.0

Le tableau suivant répertorie la version de Sqoop incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Sqoop.

Pour la version des composants installés avec Sqoop dans cette version, consultez la version [7.2.0 des versions](#) des composants.

Informations sur la version de Sqoop pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Sqoop	Composants installés avec Sqoop
emr-7.2.0	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Étiquette EMR de sortie Amazon	Version Sqoop	Composants installés avec Sqoop
		timeline-server, mariadb-server, sqoop-client

Version Sqoop pour 6.15.0

Le tableau suivant répertorie la version de Sqoop incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Sqoop.

Pour connaître la version des composants installés avec Sqoop dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Sqoop pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Sqoop	Composants installés avec Sqoop
emr-6.15.0	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Version Sqoop pour 5.36.2

Le tableau suivant répertorie la version de Sqoop incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Sqoop.

Pour la version des composants installés avec Sqoop dans cette version, voir la [version 5.36.2 Versions des composants](#).

Informations sur la version de Sqoop pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Sqoop	Composants installés avec Sqoop
emr-5.36.2	Sqoop 1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Considérations relatives à Sqoop sur Amazon EMR

Tenez compte des éléments suivants lorsque vous exécutez Sqoop sur Amazon EMR.

Utilisation de Sqoop avec intégration HCatalog

Sqoop sur Amazon EMR prend en charge l'intégration de [Sqoop](#). HCatalog Lorsque vous utilisez Sqoop pour écrire une sortie dans une HCatalog table dans Amazon S3, désactivez Amazon EMR Direct Write en définissant les `mapred.output.direct.EmrFileSystem` propriétés `mapred.output.direct.NativeS3FileSystem` et sur `false`. Pour plus d'informations, consultez [En utilisant HCatalog](#). Vous pouvez utiliser les commandes Hadoop `-D mapred.output.direct.NativeS3FileSystem=false` et `-D mapred.output.direct.EmrFileSystem=false`. Si vous ne désactivez pas l'écriture directe, aucune erreur ne se produit, mais la table est créée dans Amazon S3 et aucune donnée n'est écrite.

Support de Sqoop JDBC et de base de données

Par défaut, un pilote MariaDB et Postgre sont installés sur Sqoop. SQL Le SQL pilote Postgre installé pour Sqoop ne fonctionne que pour Postgre 8.4. SQL Pour installer un autre ensemble de JDBC connecteurs pour Sqoop, connectez-vous au nœud principal du cluster et installez-les dedans. / `usr/lib/sqoop/lib` Vous trouverez ci-dessous des liens vers différents JDBC connecteurs :

- MariaDB : [A propos de MariaDB Connector/J](#).
- Postgre SQL : pilote [Postgre SQL JDBC](#).
- SQLServer: [Téléchargez le JDBC pilote Microsoft pour SQL serveur](#).
- Mon SQL : [Download Connector/J](#)
- Oracle : [obtenir des JDBC pilotes Oracle et UCP depuis le référentiel Oracle Maven](#)

Les bases de données prises en charge pour Sqoop sont répertoriées à l'adresse suivante, http://sqoop.apache.org/docs/VERSION/SqoopUserGuide.html#_supported_databases, où *version* est la version de Sqoop que vous utilisez, par exemple 1.4.6. Si la chaîne de JDBC connexion ne correspond pas à celles de cette liste, vous devez spécifier un pilote.

Par exemple, vous pouvez exporter vers une table de base de données Amazon Redshift à l'aide de la commande suivante (pour la JDBC version 4.1) :

```
sqoop export --connect jdbc:redshift://$MYREDSHIFTHOST:5439/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver com.amazon.redshift.jdbc41.Driver --
username master --password Mymasterpass1
```

Vous pouvez utiliser à la fois les chaînes de connexion MariaDB et SQL My, mais si vous spécifiez la chaîne de connexion MariaDB, vous devez spécifier le pilote :

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

Si vous utilisez le chiffrement Secure Socket Layer pour accéder à votre base de données, vous devez utiliser un code JDBC URI similaire dans l'exemple d'exportation Sqoop suivant :

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb?
verifyServerCertificate=false&useSSL=true&requireSSL=true --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

Pour plus d'informations sur SSL le chiffrement dans RDS, consultez la section [Utilisation SSL pour chiffrer une connexion à une instance](#) de base de données dans le guide de l'RSUtilisateur Amazon.

Pour plus d'informations, consultez la documentation [Apache Sqoop](#).

Sécurisation de votre mot de passe

Différentes méthodes permettent de sécuriser votre mot de passe :

Java KeyStore

La méthode préférée chiffre le mot de passe avec un Java KeyStore (JKS), éliminant ainsi le besoin de stocker le mot de passe dans un format lisible.

1. Créez un alias de mot de passe. Lorsque vous y êtes invité, entrez le mot de passe que vous utilisez pour accéder à la base de données.

```
hadoop credential create mydb.password.alias -provider jceks://hdfs/user/root/
mysql.password.jceks
```

2. Utilisez l'alias de mot de passe pour lancer la tâche Sqoop :

```
sqoop export -Dhadoop.security.credential.provider.path=jceks://hdfs/user/
root/mysql.password.jceks --connect jdbc:mariadb://$HOSTNAME:3306/mydb
--table mysqoopexport --export-dir s3://mybucket/myinputfiles/ --driver
org.mariadb.jdbc.Driver --username master --password-alias mydb.password.alias
```

--fichier-motdepasse

Vous pouvez utiliser la commande `--password-file` pour transmettre le mot de passe dans un fichier comme illustré dans l'exemple suivant :

1. Créez un fichier contenant le mot de passe :

```
echo -n 'Mymasterpass1' > /home/hadoop/mysql-pass.password
```

2. Utilisez le fichier pour lancer la tâche Sqoop :

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password-file /home/hadoop/mysql-pass.password
```

-P

Vous pouvez utiliser la commande -P pour entrer le mot de passe dans une invite comme illustré dans l'exemple suivant :

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport --
export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --username
master -P
```

Historique des versions de Sqoop

Le tableau suivant répertorie la version de Sqoop incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section [Version des composants de votre version](#) dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Sqoop

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-7.2.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.36.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
		hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema- nager, hadoop-yarn-resour- cemanager, hadoop-yarn- timeline-server, mariadb-s- erver, sqoop-client
emr-7.1.0	1.4.7	emrfs, emr-ddb, emr-goodi- es, hadoop-client, hadoop-ma- pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema- nager, hadoop-yarn-resour- cemanager, hadoop-yarn- timeline-server, mariadb-s- erver, sqoop-client
emr-7.0.0	1.4.7	emrfs, emr-ddb, emr-goodi- es, hadoop-client, hadoop-ma- pred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop- hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodema- nager, hadoop-yarn-resour- cemanager, hadoop-yarn- timeline-server, mariadb-s- erver, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.15.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.14.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.13.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.12.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.11.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.11.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.10.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.10.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.9.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.9.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.8.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.8.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.7.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.36.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.36.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.6.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.35.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.5.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.4.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.3.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.3.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.2.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.2.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-6.1.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-6.1.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.34.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.33.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.33.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.32.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.32.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.31.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.31.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.30.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
emr-5.30.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.30.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client
mr-5.29.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.28.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.28.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.27.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.27.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.26.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.25.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.24.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.24.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.23.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.23.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.22.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.21.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.21.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.21.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.20.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.20.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.19.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.19.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.18.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.18.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.17.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.17.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.17.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.16.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.16.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.15.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.15.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.14.2	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.14.1	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.14.0	1.4.7	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.13.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.13.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.12.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.12.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.12.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.12.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.11.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.11.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.11.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.11.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.10.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.10.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.9.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.9.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.8.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.8.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.8.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.8.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.7.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.7.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.6.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.6.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mysql-server, sqoop-client
emr-5.5.4	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.5.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.5.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.4.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.4.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.3.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.3.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.3.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.2.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.2.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.2.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.2.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.1.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.1.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.3	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Étiquette Amazon EMR Release	Version Sqoop	Composants installés avec Sqoop
emr-5.0.2	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.1	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-5.0.0	1.4.6	emrfs, emr-ddb, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

TensorFlow

TensorFlow est une bibliothèque mathématique symbolique open source pour les applications d'intelligence artificielle et d'apprentissage profond. Pour plus d'informations, consultez le [TensorFlow site Web](#). TensorFlow est disponible avec les EMR versions 5.17.0 et ultérieures d'Amazon.

Le tableau suivant répertorie la version TensorFlow incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants avec lesquels Amazon EMR installe TensorFlow.

Pour la version des composants installés TensorFlow dans cette version, voir [Versions des composants de la version 7.2.0](#).

TensorFlow informations de version pour emr-7.2.0

Étiquette EMR de sortie Amazon	TensorFlow Version	Composants installés avec TensorFlow
emr-7.2.0	TensorFlow 2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Le tableau suivant répertorie la version TensorFlow incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants avec lesquels Amazon EMR installe TensorFlow.

Pour connaître la version des composants installés TensorFlow dans cette version, consultez la [version 6.15.0 Versions des composants](#).

TensorFlow informations de version pour emr-6.15.0

Étiquette EMR de sortie Amazon	TensorFlow Version	Composants installés avec TensorFlow
emr-6.15.0	TensorFlow 2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Le tableau suivant répertorie la version TensorFlow incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants avec lesquels Amazon EMR installe TensorFlow.

Pour la version des composants installés TensorFlow dans cette version, voir la [version 5.36.2 Versions des composants](#).

TensorFlow informations de version pour emr-5.36.2

Étiquette EMR de sortie Amazon	TensorFlow Version	Composants installés avec TensorFlow
emr-5.36.2	TensorFlow 2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

TensorFlow compilations par type d'EC2instance Amazon

Amazon EMR utilise différentes versions de la TensorFlow bibliothèque en fonction des types d'instances que vous choisissez pour votre cluster. Le tableau suivant répertorie les builds par type d'instance.

EC2types d'instances	TensorFlow construire
M5 et C5	Tensorflow 1.9.0 avec optimisation Intel MKL
P2	Tensorflow 1.9.0 avec CUDA 9.2, cu 7.1 DNN
P3	Tensorflow 1.9.0 avec CUDA 9.2, cu 7.1, 2.2.13 DNN NCCL Nvidia n'NCCLest disponible que sur les instances P3. Contrat de licence utilisateur final (EULA) : En utilisant les composants Nvidia sur AmazonEMR, vous acceptez les termes et conditions décrits dans le produit EULA .
Tous les autres	Tensorflow 1.9.0

Sécurité

En plus de suivre les instructions de la [section Utilisation TensorFlow sécurisée](#), nous vous recommandons de lancer votre cluster dans un sous-réseau privé afin de limiter l'accès aux sources fiables. Pour plus d'informations, consultez les [VPCoptions Amazon](#) dans le guide EMR de gestion Amazon.

En utilisant TensorBoard

TensorBoard est une suite d'outils de visualisation pour les TensorFlow programmes. Pour plus d'informations, voir [TensorBoard: Apprentissage visualisé sur le site](#) Web de Tensorflow.

Pour l'utiliser TensorBoard avec AmazonEMR, vous devez commencer par TensorBoard le nœud principal du cluster.

Pour utiliser Tensorboard avec Tensorflow sur Amazon EMR

1. Connectez-vous au nœud principal du cluster à l'aide deSSH. Pour plus d'informations, consultez [Connect to the master node SSH à l'aide](#) de l'Amazon EMR Management Guide.
2. Tapez la commande suivante pour démarrer Tensorboard sur le nœud principal. Remplacez *my/log/directory* par un répertoire sur le nœud principal où vous avez généré et résumé des données stockées à l'aide d'un enregistreur de résumé.

Amazon EMR 5.19.0 and later

```
python3 -m tensorboard.main --logdir=/home/hadoop/tensor --bind_all
```

Amazon EMR 5.18.1 and earlier

```
python3 -m tensorboard.main --logdir=/my/log/dir
```

Par défaut, le nœud maître héberge TensorBoard en utilisant le port 6006 et le DNS nom public principal. Une fois que vous avez démarré TensorBoard, la sortie de la URL ligne de commande présente les informations auxquelles vous pouvez vous connecter TensorBoard, comme indiqué dans l'exemple suivant :

```
TensorBoard 1.9.0 at http://master-public-dns-name:6006 (Press CTRL+C to quit)
```

3. Configurez l'accès aux interfaces web sur le nœud principal à partir de clients de confiance. Pour plus d'informations, consultez [Afficher les interfaces Web hébergées sur des EMR clusters Amazon](#) dans le guide EMR de gestion Amazon.
4. Ouvert TensorBoard à `http://master-public-dns-name:6006`.

TensorFlow historique des versions

Le tableau suivant répertorie la version TensorFlow incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

TensorFlow informations sur la version

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-7.2.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.36.2	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-7.1.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-7.0.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
		hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.15.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.14.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-6.13.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.12.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.11.1	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-6.11.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.10.1	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.10.0	2.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-6.9.1	2.10.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.9.0	2.10.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.8.1	2.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-6.8.0	2.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.7.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.36.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.36.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-6.6.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.35.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-6.5.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.4.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.3.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-6.3.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.2.1	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.2.0	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-6.1.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.1.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-6.0.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-6.0.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.34.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.33.1	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.33.0	2.4.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.32.1	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.32.0	2.3.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.31.1	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.31.0	2.1.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.30.2	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.30.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.30.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
mr-5.29.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.28.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.28.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.27.1	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.27.0	1.14.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.26.0	1.13.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.25.0	1.13.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.24.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.24.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.23.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.23.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.22.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.21.2	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.21.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.21.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow
emr-5.20.1	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.20.0	1.12.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.19.1	1.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.19.0	1.11.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.18.1	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.18.0	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.17.2	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Étiquette Amazon EMR Release	TensorFlow Version	Composants installés avec TensorFlow
emr-5.17.1	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow
emr-5.17.0	1.9.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow

Apache Tez

Apache Tez est un framework qui crée un graphe acyclique dirigé complexe (DAG) de tâches de traitement de données. Vous pouvez l'utiliser comme alternative à Hadoop dans certains cas MapReduce d'utilisation. Par exemple, vous pouvez exécuter des flux de travail Pig and Hive avec Hadoop MapReduce ou vous pouvez utiliser Tez comme moteur d'exécution. Pour plus d'informations, consultez <https://tez.apache.org/>. EMR Les versions 4.7.0 et supérieures d'Amazon incluent Tez.

Le tableau suivant répertorie la version de Tez incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Tez.

Pour la version des composants installés avec Tez dans cette version, voir [Versions des composants de la version 7.2.0](#).

Informations sur la version de Tez pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Tez	Composants installés avec Tez
emr-7.2.0	Tez 0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Le tableau suivant répertorie la version de Tez incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Tez.

Pour connaître la version des composants installés avec Tez dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Tez pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Tez	Composants installés avec Tez
emr-6.15.0	Tez 0.10.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Le tableau suivant répertorie la version de Tez incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Tez.

Pour la version des composants installés avec Tez dans cette version, voir la [version 5.36.2 Versions des composants](#).

Informations sur la version de Tez pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Tez	Composants installés avec Tez
emr-5.36.2	Tez 0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Rubriques

- [Création d'un cluster avec Tez](#)
- [Configuration de Tez](#)
- [Interface utilisateur web Tez](#)
- [Serveur de chronologie](#)
- [Historique des versions de Tez](#)

Création d'un cluster avec Tez

Pour installer Tez, choisissez Apache Tez comme application lorsque vous créez votre cluster.

Pour créer un cluster avec Tez installé à l'aide de la console

1. Accédez à la nouvelle EMR console Amazon et sélectionnez Basculer vers l'ancienne console dans la navigation latérale. Pour plus d'informations sur ce qu'implique le passage à l'ancienne console, consultez la rubrique [Utilisation de l'ancienne console](#).
2. Choisissez Créer un cluster et Go to advanced options (Aller aux options avancées).
3. Dans Configuration des logiciels, choisissez une version de emr-4.7.0 ou supérieure.
4. Sélectionnez Tez ainsi que les autres applications que vous souhaitez EMR qu'Amazon installe.
5. Sélectionnez d'autres options si nécessaire, puis choisissez Create cluster (Créer le cluster).

Pour créer un cluster avec Tez à l'aide du AWS CLI

- Utilisez la commande `create-cluster` avec l'option `--applications` pour spécifier Tez. L'exemple suivant crée un cluster avec Tez installé.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (^).

```
aws emr create-cluster --name "Cluster with Tez" --release-label emr-7.2.0 \  
--applications Name=Tez --ec2-attributes KeyName=myKey \  

```

```
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Configuration de Tez

Vous pouvez personnaliser Tez en définissant des valeurs à l'aide de la classification de configuration `tez-site`, qui configure les paramètres dans le fichier de configuration `tez-site.xml`. Pour plus d'informations, consultez [TezConfiguration](#) la documentation d'Apache Tez. Pour modifier Hive ou Pig afin d'utiliser le moteur d'exécution Tez, utilisez les classifications de configuration `hive-site` et `pig-properties` correspondantes. Des exemples sont présentés ci-après.

Exemple de configuration

Exemple Exemple : Personnalisation du niveau d'enregistrement racine de Tez et Configuration de Tez comme moteur d'exécution pour Hive et Pig

L'exemple de commande `create-cluster` ci-dessous crée un cluster avec Tez, Hive et Pig installés. La commande fait référence à un fichier stocké dans Amazon S3, `myConfig.json`, qui spécifie les propriétés pour la classification `tez-site` qui définit `tez.am.log.level` sur `DEBUG` et définit le moteur d'exécution Tez pour Hive et Pig en utilisant les classifications de configuration `hive-site` et `pig-properties`.

Note

Les caractères de continuation de ligne Linux (`\`) sont inclus pour des raisons de lisibilité. Ils peuvent être supprimés ou utilisés dans les commandes Linux. Pour Windows, supprimez-les ou remplacez-les par un caret (`^`).

```
aws emr create-cluster --release-label emr-7.2.0 \  
--applications Name=Tez Name=Hive Name=Pig --ec2-attributes KeyName=myKey \  
--instance-type m5.xlarge --instance-count 3 \  
--configurations https://s3.amazonaws.com/mybucket/myfolder/myConfig.json --use-  
default-roles
```

Vous trouverez ci-dessous des exemples de contenus `myConfig.json`.

```
[
```

```
{
  "Classification": "tez-site",
  "Properties": {
    "tez.am.log.level": "DEBUG"
  }
},
{
  "Classification": "hive-site",
  "Properties": {
    "hive.execution.engine": "tez"
  }
},
{
  "Classification": "pig-properties",
  "Properties": {
    "exectype": "tez"
  }
}
]
```

Note

Avec Amazon EMR version 5.21.0 et versions ultérieures, vous pouvez remplacer les configurations de cluster et spécifier des classifications de configuration supplémentaires pour chaque groupe d'instances d'un cluster en cours d'exécution. Pour ce faire, utilisez la EMR console Amazon, le AWS Command Line Interface (AWS CLI) ou le AWS SDK. Pour plus d'informations, consultez [Fourniture d'une configuration pour un groupe d'instances dans un cluster en cours d'exécution](#).

Ouverture fractionnée asynchrone Tez

Lorsque le chemin de la table contient un grand nombre de petits fichiers et qu'une requête tente de tous les lire, chaque petit fichier correspondant à chaque division est combiné en une unité de fractionnement groupée Tez. Un seul mappeur traite ensuite la division groupée unique Tez. Puisque l'exécution est synchrone, chaque division de la division groupée est traitée une par une. Les objets `RecordReader` doivent donc traiter les divisions de manière synchrone.

Nom	Classification	Description
<code>tez.grouping.split</code> <code>.init.threads</code>	<code>tez-site</code>	Spécifie le nombre de threads de démon que Tez utilise pour pré-initialiser les <code>RecordReaders</code> et les divisions ouvertes. Pour les tables ACID, la valeur maximale prise en charge pour <code>tez.grouping.split</code> <code>.init.threads</code> est 1.
<code>tez.grouping.split</code> <code>.init.recordreaders</code>	<code>tez-site</code>	Spécifie le nombre de <code>RecordReaders</code> dont la pré-initialisation doit être conservée par les threads de démon. Cela peut être utile lorsque la division groupée Tez contient un grand nombre de <code>InputSplits</code> . L'initialisation de <code>RecordReaders</code> pour traiter ces divisions d'entrée peut être effectuée de manière asynchrone avec des threads de démon au lieu d'un traitement séquentiel.

Analyse comparative de l'ouverture fractionnée asynchrone Tez

Nous avons utilisé les environnements et configurations suivants pour procéder à l'analyse comparative de la capacité d'ouverture fractionnée asynchrone Tez :

- Environnement de référence : EMR cluster Amazon avec 1 nœud principal qui utilise `m5.16xlarge` et 16 nœuds principaux qui utilisent `m5.16xlarge`.

- Configurations d'évaluation – Pour simuler le scénario d'analyse comparative dans lequel un grand nombre de divisions d'entrée se trouvent dans une seule division groupée Tez, la valeur `tez.grouping.split-count` est définie sur 1.
- Table utilisée pour l'analyse comparative – La table contient 200 partitions, chaque partition contenant un seul fichier. L'analyse comparative est effectuée lorsque cette table contient des fichiers CSV et lorsque cette table contient des fichiers parquet. Requête Hive pour l'analyse comparative : `SELECT COUNT(*)` à dix reprises dans la table, et prenez le temps d'exécution moyen.
- Configurations pour activer l'ouverture fractionnée asynchrone Tez :
 - `tez.grouping.split.init.threads = 4`
 - `tez.grouping.split.init.recordreaders = 10`

Jeux de données	Fonctionnalité désactivée (référence)	Fonctionnalité activée	Amélioration
Jeu de données CSV	90,26 secondes	79,20 secondes	12,25 %
Jeu de données Parquet	54,67 secondes	42,23 secondes	22,75 %

Interface utilisateur web Tez

Tez a sa propre interface utilisateur web. Pour afficher l'interface utilisateur Web, consultez ce qui suitURL.

```
http://masterDNS:8080/tez-ui
```

Pour activer l'onglet Hive Queries sur l'interface utilisateur Web de Tez, définissez la configuration suivante.

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.exec.pre.hooks": "org.apache.hadoop.hive.q1.hooks.ATSHook",
      "hive.exec.post.hooks": "org.apache.hadoop.hive.q1.hooks.ATSHook",
```



```

    "hive.exec.failure.hooks": "org.apache.hadoop.hive.q1.hooks.ATSHook"
  }
}
]

```

Vous pouvez également consulter les détails de Tez, de Spark et de YARN l'interface utilisateur de l'application à l'aide des liens figurant dans l'onglet Interfaces utilisateur de l'application de la page détaillée d'un cluster dans la console. Les interfaces utilisateur (UI) des EMR applications Amazon sont hébergées hors cluster et sont disponibles après la fin du cluster. Ils ne vous obligent pas à configurer une SSH connexion ou un proxy Web, ce qui vous permet de résoudre et d'analyser plus facilement les tâches actives et l'historique des tâches.

Pour plus d'informations, consultez la section [Afficher l'historique des applications](#) dans le guide EMR de gestion Amazon.

Serveur de chronologie

Le serveur YARN Timeline est configuré pour fonctionner lorsque Tez est installé. Pour afficher les tâches soumises via Tez ou des moteurs MapReduce d'exécution à l'aide du serveur Timeline, consultez l'interface utilisateur Web à l'aide du URL `http://master-public-DNS:8188`. Pour plus d'informations, consultez [Afficher les interfaces Web hébergées sur des EMR clusters Amazon](#) dans le guide EMR de gestion Amazon.

Historique des versions de Tez

Le tableau suivant répertorie la version de Tez incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Tez

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-7.2.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
		hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-5.36.2	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-7.1.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-7.0.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.15.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.14.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-6.13.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.12.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.11.1	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-6.11.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.10.1	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker
emr-6.10.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-6.9.1	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.9.0	0,1,2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.8.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-6.8.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.7.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.36.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.36.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.6.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.35.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-6.5.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.4.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.3.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-6.3.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.2.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.2.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-6.1.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.1.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-6.0.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-6.0.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.34.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.33.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.33.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.32.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.32.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.31.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.31.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.30.2	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.30.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.30.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
mr-5.29.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.28.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.28.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.27.1	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.27.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.26.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.25.0	0.9.2	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.24.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.24.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.23.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.23.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.22.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.21.2	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.21.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.21.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.20.1	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.20.0	0.9.1	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.19.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.19.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.18.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.18.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.17.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.17.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.17.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.16.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.16.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.15.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.15.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.14.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.14.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.14.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.13.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.13.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.12.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.12.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.12.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.12.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.11.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.11.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.11.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.10.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.10.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.9.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.9.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.8.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.8.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.8.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.8.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.7.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.7.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.6.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.6.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.5.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.5.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.5.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.4.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.4.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.3.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.3.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.3.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.2.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.2.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.2.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.1.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.1.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.0.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.0.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-5.0.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-5.0.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.6	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.5	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-4.9.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.9.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-4.9.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.5	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.4	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-4.8.3	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.2	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.8.1	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-4.8.0	0.8.4	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.4	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.3	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Étiquette Amazon EMR Release	Version Tez	Composants installés avec Tez
emr-4.7.2	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.1	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn
emr-4.7.0	0.8.3	emrfs, emr-goodies, hadoop-client, hadoop-mapred, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tez-on-yarn

Notes de publication de Tez par version

Rubriques

- [Amazon EMR 7.1.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.15.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.14.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.13.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.12.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.11.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.10.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.9.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.8.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.7.0 - Notes de mise à jour de Tez](#)
- [Amazon EMR 6.6.0 - Notes de mise à jour de Tez](#)

Amazon EMR 7.1.0 - Notes de mise à jour de Tez

Amazon EMR 7.1.0 - Essayez les modifications

Type	Description
Correctif de bogue.	TEZ-4394 — Netty4 ShuffleHandler : devrait utiliser 1 thread de boss (#256)
Upgrade	Met à jour Netty vers la version 4.1.100.Final.
Upgrade	Met à jour Jetty vers la version 9.4.53.v20231009.

Amazon EMR 6.15.0 - Notes de mise à jour de Tez

Amazon EMR 6.15.0 - Tez les modifications

Type	Description
Fonctionnalité	TEZ-4397 : Open Tez Input se divise de manière asynchrone

Type	Description
Upgrade	TEZ-4493 : mise à niveau d'Apache Hadoop vers la version 3.3.6

Amazon EMR 6.15.0 - Fonctionnalités de Tez

- [Ouverture de division asynchrone Tez](#) — Amazon EMR 6.15.0 introduit des configurations que vous pouvez spécifier pour ouvrir de manière asynchrone les divisions d'entrée dans une division groupée Tez. La fonctionnalité a été lancée par [TEZ-4397](#), mais présentait des régressions dans Hive. OSS Amazon EMR Hive a corrigé les régressions et les bogues supplémentaires dans le tableau Hive. ACID Cette amélioration permet d'accélérer les performances des requêtes de lecture lorsqu'il existe un grand nombre de divisions d'entrée dans une seule division groupée Tez. Pour plus d'informations, consultez [Ouverture fractionnée asynchrone Tez](#).

Amazon EMR 6.14.0 - Notes de mise à jour de Tez

Amazon EMR 6.14.0 - Tez les modifications

Type	Description
Amélioration	Mise à niveau de TLS la version dans Tez vers la version 1.2

Amazon EMR 6.13.0 - Notes de mise à jour de Tez

Amazon EMR 6.13.0 - Tez les modifications

Type	Description
Correctif de bogue.	REVERT TEZ-4295 : Impossible de décompresser les données. La longueur de la mémoire tampon est trop petite.

Type	Description
Correctif de bogue.	REVERT TEZ-4302 : compatible avec <code>NullPointerException</code> . <code>CodecUtils</code> <code>GzipCodec</code>
Correctif de bogue.	REVERT TEZ-4234 : Le compresseur peut provoquer une situation <code>Buffer.limit</code> où <code>IllegalArgumentException</code> la limite dépasse la capacité.
Correctif de bogue.	REVERT TEZ-4135 : Améliore l'allocation de mémoire lors de l'exécution de lectures en mémoire.

Amazon EMR 6.12.0 - Notes de mise à jour de Tez

Amazon EMR 6.12.0 - Tez les modifications

Type	Description
Amélioration	Support ajouté pour les JDK environnements d'exécution 11 et JDK 17
Correctif de bogue.	TEZ-4492 : Mettre à jour <code>Bowerrc</code> pour utiliser le miroir <code>bower.herokuapp</code> afin d'éviter le problème <code>Bower Registry _ (-2608) CERT EXPIRE BOWER</code>
Upgrade	Mise à niveau de <code>Surefire</code> vers la version 3.0.0-M7

Amazon EMR 6.11.0 - Notes de mise à jour de Tez

Amazon EMR 6.11.0 - Tez les modifications

Type	Description
Bogue	Correction d'une transition d'état de sommet non valide lors du nettoyage des données réorganisées au niveau des sommets
Bogue	Le nettoyage des données de shuffle au niveau fixe DAG ou au niveau du sommet ne fonctionne pas
Amélioration	Activez <code>tez.am.dag.cleanup.on.completion</code> par défaut pour effacer les données de shuffle une fois terminé DAGs

Amazon EMR 6.10.0 - Notes de mise à jour de Tez

Amazon EMR 6.10.0 - Tez les modifications

Type	Description
Fonctionnalité	Activation de <code>tez.runtime.transfer.data-via-events.enabled</code> par défaut
Rétroportage	TEZ-4450 : correction de l'échec de récupération des données de shuffle lorsque les données de shuffle sont transférées via des événements de déplacement de données
Rétroportage	TEZ-4460 : correction d'une erreur de délai de lecture lors de la récupération des données de shuffle depuis le gestionnaire Tez Shuffle

Type	Description
Rétroportage	TEZ-4455 : Ajout d'un ShuffleHandler pipeline pour une LoggingHandler meilleure débuggabilité
Bogue	Corrige le blocage de la tâche Tez par intermittence lorsque la préemption de la tâche est activée

Amazon EMR 6.9.0 - Notes de mise à jour de Tez

Amazon EMR 6.9.0 - Essayez les modifications

Type	Description
Upgrade	Tez est mis à niveau vers la version 0.10.2. Pour plus d'informations, consultez le journal des modifications pour Apache Tez 0.10.2 .
Upgrade	Mise à niveau de Hadoop vers la version 3.3.3.
Bogue	Désactivé <code>tez.runtime.transfer.data-via-events.enabled</code> par défaut en raison de TEZ-4450 .

Amazon EMR 6.8.0 - Notes de mise à jour de Tez

Amazon EMR 6.8.0 - Essayez les modifications

Type	Description
Rétroportage	TEZ-3363 : Supprimer les données intermédiaires au niveau du sommet pour Shuffle Handler

Type	Description
Rétroportage	TEZ-4129 : Supprimer les données de tentative intermédiaires en cas d'échec des tentatives pour Shuffle Handler
Rétroportage	TEZ-4430 : Correction de la propriété tez.task.launch.cmd-opts qui ne fonctionnait pas

Amazon EMR 6.7.0 - Notes de mise à jour de Tez

Amazon EMR 6.7.0 - Essayez les modifications

Type	Description
Rétroportage	TEZ-4403 : Mise à niveau de la version vers la SLF4J version 1.7.36
Rétroportage	TEZ-4405 : Remplacer log4j 1.x par reload4j
Rétroportage	TEZ-4411 : Échec de la compilation de Tez : FileSaver .js introuvable

Amazon EMR 6.6.0 - Notes de mise à jour de Tez

Amazon EMR 6.6.0 - Essayez les modifications

Type	Description
Rétroportage	TEZ-3918 : correction de la propriété tez.task.log.level qui ne fonctionnait pas.
Rétroportage	TEZ-4353 : Mise à jour de commons-io vers la version 2.8.0.
Rétroportage	TEZ-4114 : Supprime la dépendance directe à la jetée de tez.

Type	Description
Rétroportage	TEZ-4323 : Les pots Jetty ont été retirés du package dist avec -4114. TEZ

Apache Zeppelin

Utilisez Apache Zeppelin sous forme de bloc-notes pour l'exploration des données interactives. Pour plus d'informations sur Zeppelin, consultez <https://zeppelin.apache.org/>. Zeppelin est inclus dans les EMR versions 5.0.0 et ultérieures d'Amazon. Les versions antérieures incluent Zeppelin en tant qu'application sandbox. Pour plus d'informations, consultez [Versions publiées par Amazon EMR 4.x](#).

Pour accéder à l'interface Web Zeppelin, configurez un SSH tunnel vers le nœud principal et une connexion proxy. Pour plus d'informations, voir [Afficher les interfaces Web hébergées sur EMR des clusters](#).

Le tableau suivant répertorie la version de Zeppelin incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants qu'Amazon EMR installe avec Zeppelin.

Pour la version des composants installés avec Zeppelin dans cette version, voir la [version 7.2.0](#) Versions des composants.

Informations sur la version de Zeppelin pour emr-7.2.0

Étiquette EMR de sortie Amazon	Version Zeppelin	Composants installés avec Zeppelin
emr-7.2.0	Zeppelin 0.10.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Le tableau suivant répertorie la version de Zeppelin incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants qu'Amazon EMR installe avec Zeppelin.

Pour connaître la version des composants installés avec Zeppelin dans cette version, consultez les [versions des composants de la version 6.15.0](#).

Informations sur la version de Zeppelin pour emr-6.15.0

Étiquette EMR de sortie Amazon	Version Zeppelin	Composants installés avec Zeppelin
emr-6.15.0	Zeppelin 0.10.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Le tableau suivant répertorie la version de Zeppelin incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants qu'Amazon EMR installe avec Zeppelin.

Pour la version des composants installés avec Zeppelin dans cette version, voir la [version 5.36.2](#) Versions des composants.

Informations sur la version de Zeppelin pour emr-5.36.2

Étiquette EMR de sortie Amazon	Version Zeppelin	Composants installés avec Zeppelin
emr-5.36.2	Zeppelin 0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-

Étiquette EMR de sortie Amazon	Version Zeppelin	Composants installés avec Zeppelin
		httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Rubriques

- [Considérations relatives à l'utilisation de Zeppelin sur Amazon EMR](#)
- [Historique des versions de Zeppelin](#)

Considérations relatives à l'utilisation de Zeppelin sur Amazon EMR

- Connectez-vous à Zeppelin en utilisant la même [méthode de SSH tunneling](#) pour vous connecter à d'autres serveurs Web sur le nœud principal. Le serveur Zeppelin est disponible sur le port 8890.
- [Les EMR versions 5.0.0 et ultérieures de Zeppelin sur Amazon prennent en charge l'authentification Shiro.](#)
- Les EMR versions 5.8.0 et ultérieures de Zeppelin sur Amazon prennent en charge l'utilisation de AWS Glue Data Catalog comme métastore pour Spark. SQL Pour plus d'informations, consultez [Utiliser le catalogue de données AWS Glue comme métastore pour Spark. SQL](#)
- Zeppelin n'utilise pas certains des paramètres définis dans le fichier de spark-defaults.conf configuration de votre cluster, même s'il indique d'allouer des exécuteurs YARN de manière dynamique si vous l'avez défini. spark.dynamicAllocation.enabled true Vous devez définir des paramètres de l'exécuteur (comme la mémoire et les noyaux) en cliquant sur l'onglet Interpreter (Interprète) de Zeppelin, puis redémarrer l'interprète pour en permettre l'utilisation.
- EMR Les versions 6.10.0 et supérieures d'Amazon prennent en charge l'intégration d'Apache Zeppelin avec Apache Flink. Pour plus d'informations, consultez [Travailler avec les jobs Flink de Zeppelin dans Amazon EMR.](#)

- Zeppelin sur Amazon EMR ne prend pas en charge l'interpréteur SparkR.

Historique des versions de Zeppelin

Le tableau suivant répertorie la version de Zeppelin incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

Informations sur la version de Zeppelin

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-7.2.0	0,1,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.36.2	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
		timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-7.1.0	0,1,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-7.0.0	0,1,1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.15.0	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.14.0	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.13.0	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.12.0	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.11.1	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.11.0	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.10.1	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.10.0	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.9.1	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.9.0	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.8.1	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.8.0	0,1,1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.7.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.36.1	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.36.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.6.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.35.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.5.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.4.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.3.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.3.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.2.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.2.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.1.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.1.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-6.0.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-6.0.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.34.0	0.10.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.33.1	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.33.0	0.9.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.32.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.32.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.31.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.31.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.30.2	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.30.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.30.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
mr-5.29.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.28.1	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.28.0	0.8.2	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.27.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.27.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.26.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.25.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.24.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.24.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.23.1	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.23.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.22.0	0.8.1	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.21.2	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.21.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.21.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.20.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.20.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.19.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.19.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.18.1	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.18.0	0.8.0	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.17.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.17.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.17.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.16.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.16.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.15.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.15.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.14.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.14.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.14.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.13.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.13.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.12.3	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.12.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.12.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.12.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.11.4	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.11.3	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.11.2	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.11.1	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.11.0	0.7.3	aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.10.1	0.7.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.10.0	0.7.3	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.9.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.9.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.8.3	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.8.2	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.8.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.8.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.7.1	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.7.0	0.7.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.6.1	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.6.0	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.4	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.5.3	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.2	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.5.1	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.5.0	0.7.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.4.1	0.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.4.0	0.7.0	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.3.2	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.3.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.3.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.2.3	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.2.2	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.2.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.2.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.1.1	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.1.0	0.6.2	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.0.3	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.0.2	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server
emr-5.0.1	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Étiquette Amazon EMR Release	Version Zeppelin	Composants installés avec Zeppelin
emr-5.0.0	0.6.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Apache ZooKeeper

Apache ZooKeeper est un service centralisé permettant de gérer les informations de configuration, de nommer, de fournir une synchronisation distribuée et de fournir des services de groupe. Pour plus d'informations sur ZooKeeper, consultez <http://zookeeper.apache.org/>.

Le tableau suivant répertorie la version ZooKeeper incluse dans la dernière version de la série Amazon EMR 7.x, ainsi que les composants avec lesquels Amazon EMR installe. ZooKeeper

Pour la version des composants installés ZooKeeper dans cette version, voir [Versions des composants de la version 7.2.0](#).

ZooKeeper informations de version pour emr-7.2.0

Étiquette EMR de sortie Amazon	ZooKeeper Version	Composants installés avec ZooKeeper
emr-7.2.0	ZooKeeper 3.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Le tableau suivant répertorie la version ZooKeeper incluse dans la dernière version de la série Amazon EMR 6.x, ainsi que les composants avec lesquels Amazon EMR installe. ZooKeeper

Pour connaître la version des composants installés ZooKeeper dans cette version, consultez la [version 6.15.0 Versions des composants](#).

ZooKeeper informations de version pour emr-6.15.0

Étiquette EMR de sortie Amazon	ZooKeeper Version	Composants installés avec ZooKeeper
emr-6.15.0	ZooKeeper 3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Le tableau suivant répertorie la version ZooKeeper incluse dans la dernière version de la série Amazon EMR 5.x, ainsi que les composants avec lesquels Amazon EMR installe ZooKeeper

Pour la version des composants installés ZooKeeper dans cette version, voir la [version 5.36.2 Versions des composants](#).

ZooKeeper informations de version pour emr-5.36.2

Étiquette EMR de sortie Amazon	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.36.2	ZooKeeper 3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Rubriques

- [ZooKeeper historique des versions](#)

ZooKeeper historique des versions

Le tableau suivant répertorie la version ZooKeeper incluse dans chaque version d'AmazonEMR, ainsi que les composants installés avec l'application. Pour connaître les versions des composants de chaque version, consultez la section Version des composants de votre version dans [Versions publiées par Amazon EMR 7.x](#), [Versions publiées par Amazon EMR 6.x](#) ou [Versions publiées par Amazon EMR 5.x](#).

ZooKeeper informations sur la version

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-7.2.0	3.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.36.2	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-7.1.0	3.9.1	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-7.0.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.15.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-6.14.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.13.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.12.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-6.11.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.11.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.10.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-6.10.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.9.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.9.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-6.8.1	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.8.0	3.5.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.7.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.36.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.36.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.6.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.35.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.5.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.4.0	3.5.7	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-6.3.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.3.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.2.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-6.2.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.1.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.1.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-6.0.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-6.0.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.34.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.33.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.33.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.32.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.32.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.31.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.31.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.30.2	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.30.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.30.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
mr-5.29.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.28.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.28.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.27.1	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.27.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.26.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.25.0	3.4.14	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.24.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.24.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.23.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.23.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.22.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.21.2	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.21.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.21.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.20.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.20.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.19.1	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.19.0	3.4.13	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.18.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.18.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.17.2	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.17.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.17.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.16.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.16.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.15.1	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.15.0	3.4.12	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.14.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.14.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.14.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.13.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.13.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.12.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.12.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.12.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.12.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.4	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.11.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.11.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.10.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.10.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.9.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.9.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.8.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.8.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.7.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.7.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.6.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server
emr-5.6.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.5.4	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.5.3	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.5.2	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.5.1	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.5.0	3.4.10	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.4.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.4.0	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.3.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.3.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.3.0	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.2.3	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.2.2	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.2.1	3.4.9	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.2.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.1.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.1.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.0.3	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server
emr-5.0.2	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server

Étiquette Amazon EMR Release	ZooKeeper Version	Composants installés avec ZooKeeper
emr-5.0.1	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server
emr-5.0.0	3.4.8	emrfs, emr-goodies, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, zookeeper-client, zookeeper-server

Connecteurs et utilitaires

Amazon EMR fournit plusieurs connecteurs et utilitaires permettant d'accéder à d'autres AWS services en tant que sources de données. Vous pouvez généralement accéder aux données de ces services au sein d'un programme. Par exemple, vous pouvez spécifier un flux Kinesis dans une requête Hive, un script Pig ou une MapReduce application, puis opérer sur ces données.

Rubriques

- [Exportez, importez, interrogez et joignez des tables dans DynamoDB à l'aide d'Amazon EMR](#)
- [Kinesis](#)
- [S3 DistCp \(s3-dist-cp\)](#)
- [Nettoyage après l'échec des DistCp tâches S3](#)

Exportez, importez, interrogez et joignez des tables dans DynamoDB à l'aide d'Amazon EMR

Note

Le connecteur Amazon EMR -DynamoDB est open source sur GitHub. Pour plus d'informations, consultez <https://github.com/aws-labs/emr-dynamodb-connector>.


DynamoDB est un service SQL sans base de données entièrement géré qui fournit des performances rapides et prévisibles ainsi qu'une évolutivité sans faille. Les développeurs peuvent créer une table de base de données et développer sans limite son trafic de demandes ou son stockage. DynamoDB répartit automatiquement les données et le trafic de la table sur un nombre de serveurs approprié afin de gérer le volume de requêtes spécifié par le client et la quantité de données stockées, tout en assurant la cohérence et la rapidité des performances. Amazon EMR et Hive vous permettent de traiter rapidement et efficacement de grandes quantités de données, telles que les données stockées dans DynamoDB. Pour plus d'informations sur DynamoDB, consultez le [manuel du développeur Amazon DynamoDB](#).

Apache Hive est une couche logicielle que vous pouvez utiliser pour interroger, cartographier et réduire les clusters à l'aide d'un langage de requête SQL simplifié appelé HiveQL. Il s'exécute sur

l'architecture Hadoop. Pour plus d'informations sur Hive et HiveQL, consultez le [Manuel du langage HiveQL](#). Pour plus d'informations sur Hive et AmazonEMR, consultez [Apache Hive](#).

Vous pouvez utiliser Amazon EMR avec une version personnalisée de Hive qui inclut la connectivité à DynamoDB pour effectuer des opérations sur les données stockées dans DynamoDB :

- Charger des données DynamoDB dans le système de fichiers distribué Hadoop HDFS () et les utiliser comme entrées dans un cluster Amazon. EMR
- Interrogation de données DynamoDB en direct à l'aide d'instructions SQL like (HiveQL).
- Jointure de données stockées dans DynamoDB et exportation ou interrogation par rapport aux données jointes.
- Exportation de données stockées dans DynamoDB vers Amazon S3.
- Importation de données stockées dans Amazon S3 vers DynamoDB.


 Note

[Le connecteur Amazon EMR -DynamoDB ne prend pas en charge les clusters configurés pour utiliser l'authentification Kerberos.](#)

Pour effectuer chacune des tâches suivantes, vous allez lancer un EMR cluster Amazon, spécifier l'emplacement des données dans DynamoDB et exécuter des commandes Hive pour manipuler les données dans DynamoDB.

Il existe plusieurs manières de lancer un EMR cluster Amazon : vous pouvez utiliser la EMR console Amazon, l'interface de ligne de commande (CLI), ou vous pouvez programmer votre cluster à l'aide d'un AWS SDK ou d'Amazon EMRAPI. Vous pouvez également choisir d'exécuter un cluster Hive de façon interactive ou à partir d'un script. Dans cette section, nous allons vous montrer comment lancer un cluster Hive interactif depuis la EMR console Amazon et leCLI.

L'utilisation interactive de Hive est un excellent moyen de tester les performances des requêtes et de régler votre application. Après avoir défini un ensemble de commandes Hive qui s'exécuteront régulièrement, pensez à créer un script Hive qu'Amazon EMR pourra exécuter pour vous.

 Warning

Les opérations de EMR lecture ou d'écriture d'Amazon sur une table DynamoDB sont comptabilisées dans votre débit provisionné établi, ce qui peut augmenter la fréquence

des exceptions de débit provisionné. Pour les demandes volumineuses, Amazon EMR met en œuvre de nouvelles tentatives avec un retard exponentiel afin de gérer la charge des demandes sur la table DynamoDB. L'exécution de EMR tâches Amazon en même temps que d'autres types de trafic peut vous faire dépasser le niveau de débit alloué. Vous pouvez contrôler cela en vérifiant la `ThrottleRequests` métrique sur Amazon CloudWatch. Si la charge de demandes est trop élevée, vous pouvez relancer le cluster et définir la valeur [Paramètre de pourcentage de lecture](#) ou [Paramètre de pourcentage d'écriture](#) sur une valeur inférieure pour limiter les opérations Amazon. EMR Pour de plus amples informations sur les paramètres de débit DynamoDB, veuillez consulter [Débit alloué](#).

Si une table est configurée pour le [mode à la demande](#), vous devez la remettre en mode provisionné avant d'exécuter une opération d'exportation ou d'importation. Les pipelines ont besoin d'un ratio de débit pour calculer les ressources à utiliser à partir d'un D.ynamoDBtable Le mode à la demande supprime le débit provisionné. Pour provisionner la capacité de débit, vous pouvez utiliser les métriques Amazon CloudWatch Events pour évaluer le débit agrégé utilisé par une table.

Rubriques

- [Configuration d'une table Hive pour exécuter des commandes Hive](#)
- [Exemples de commandes Hive pour l'exportation, l'importation et l'interrogation de données dans DynamoDB](#)
- [Optimisation des performances pour les EMR opérations Amazon dans DynamoDB](#)

Configuration d'une table Hive pour exécuter des commandes Hive

Apache Hive est une application d'entrepôt de données que vous pouvez utiliser pour interroger les données contenues dans les EMR clusters Amazon à l'aide d'un langage SQL similaire. Pour plus d'informations sur Hive, consultez le site <http://hive.apache.org/>.

La procédure suivante suppose que vous avez déjà créé un cluster et que vous avez spécifié une paire de EC2 clés Amazon. Pour savoir comment commencer à créer des clusters, consultez [Getting started with Amazon EMR](#) dans le Amazon EMR Management Guide.

Configurer Hive pour utiliser MapReduce

Lorsque vous utilisez Hive sur Amazon EMR pour interroger des tables DynamoDB, des erreurs peuvent se produire si Hive utilise le moteur d'exécution par défaut, Tez. C'est pourquoi, lorsque vous

créez un cluster avec Hive qui s'intègre à DynamoDB comme décrit dans cette section, nous vous recommandons d'utiliser une classification de configuration qui définit Hive à utiliser. MapReduce Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

L'extrait suivant montre la classification de configuration et les propriétés à utiliser pour définir MapReduce comme moteur d'exécution pour Hive :

```
[
    {
        "Classification": "hive-site",
        "Properties": {
            "hive.execution.engine": "mr"
        }
    }
]
```

Pour exécuter les commandes Hive de façon interactive

1. Connectez-vous au nœud maître. Pour plus d'informations, consultez [Connect to the master node SSH à l'aide](#) de l'Amazon EMR Management Guide.
2. A l'invite de commande pour le nœud maître actuel, tapez `hive`.

Vous devez voir une invite Hive : `hive>`

3. Entrez une commande Hive qui mappe une table de l'application Hive aux données dans DynamoDB. Ce tableau agit comme référence pour les données stockées dans Amazon DynamoDB ; les données ne sont pas stockées localement dans Hive et les requêtes utilisant cette table s'exécutent sur les données en direct dans DynamoDB et consomment la capacité en lecture ou en écriture de la table chaque fois qu'une commande est exécutée. Si vous prévoyez d'exécuter plusieurs commandes Hive sur le même ensemble de données, pensez d'abord à l'exporter.

L'exemple suivant montre la syntaxe pour mapper une table Hive à une table DynamoDB.

```
CREATE EXTERNAL TABLE hive_tablename
(hive_column1_name column1_datatype, hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
"hive_column1_name:dynamodb_attribute1_name, hive_column2_name:dynamodb_attribute2_name...")
```

Lorsque vous créez une table dans Hive depuis DynamoDB, vous devez la créer en tant que table externe à l'aide du mot clé EXTERNAL. La différence entre les tables internes et externes est que les données des tables internes sont supprimées lorsqu'une table interne est supprimée. Il ne s'agit pas du comportement souhaité lorsque vous êtes connecté à Amazon DynamoDB, si bien que seules les tables externes sont prises en charge.

Par exemple, la commande Hive suivante crée une table nommée `hivetable1` dans Hive qui fait référence à la table DynamoDB nommée `dynamodhtable1`. La table DynamoDB `dynamodhtable1` possède un schéma de clé primaire. L'élément de clé de hachage est `name` (type de chaîne), l'élément de clé de plage est `year` (type numérique) et chaque élément a une valeur d'attribut pour `holidays` (type string set).

```
CREATE EXTERNAL TABLE hivetable1 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");
```

La ligne 1 utilise l'instruction HiveQL `CREATE EXTERNAL TABLE`. Pour `hivetable1`, vous devez établir une colonne pour chaque paire nom-valeur d'attribut de la table DynamoDB et fournir le type de données. Ces valeurs ne sont pas sensibles à la casse et vous pouvez donner aux colonnes le nom de votre choix (à l'exception des mots réservés).

La ligne 2 utilise l'instruction `STORED BY`. La valeur de `STORED BY` est le nom de la classe qui gère la connexion entre Hive et DynamoDB. Cette valeur doit être définie sur `'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'`.

La ligne 3 utilise l'instruction `TBLPROPERTIES` pour associer « `hivetable1` » à la table et au schéma appropriés dans DynamoDB. Fournissez à `TBLPROPERTIES` les valeurs des paramètres `dynamodb.table.name` et `dynamodb.column.mapping`. Ces valeurs sont sensibles à la casse.

Note

Tous les noms d'attribut DynamoDB de la table doivent avoir les colonnes correspondantes dans la table Hive. En fonction de votre EMR version d'Amazon, les scénarios suivants se produisent si le one-to-one mappage n'existe pas :

- Sur Amazon EMR version 5.27.0 et versions ultérieures, le connecteur dispose de validations qui garantissent un mappage one-to-one entre les noms d'attributs DynamoDB et les colonnes de la table Hive. Une erreur se produira si le one-to-one mappage n'existe pas.
- Sur les EMR versions 5.26.0 et antérieures d'Amazon, la table Hive ne contiendra pas la paire nom-valeur de DynamoDB. Si vous ne mappez pas les attributs de clé primaire DynamoDB, Hive génère une erreur. Si vous ne mappez pas un attribut de clé non primaire, aucune erreur n'est générée, mais vous ne voyez pas les données de la table Hive. Si les types de données ne correspondent pas, la valeur est null.

Ensuite, vous pouvez commencer à exécuter les opérations Hive sur `hivetable1`. Les requêtes exécutées sur `hivetable1` sont exécutées en interne sur la table DynamoDB `dynamodhtable1` de votre compte DynamoDB et consomment des unités de lecture ou d'écriture à chaque exécution.

Lorsque vous exécutez les requêtes Hive sur une table DynamoDB, vous devez vous assurer que vous avez alloué une quantité suffisante d'unités de capacité de lecture.

Par exemple, supposons que vous ayez provisionné 100 unités de capacité en lecture pour votre table DynamoDB. Vous pourrez ainsi effectuer 100 lectures, ou 409 600 octets, par seconde. Si cette table contient 20 Go de données (21 474 836 480 octets) et que votre requête Hive exécute une analyse complète de la table, vous pouvez estimer la durée d'exécution de la requête :

$$21\,474\,836\,480 / 409\,600 = 52\,429 \text{ secondes} = 14,56 \text{ heures}$$

Le seul moyen de diminuer le temps nécessaire consiste à ajuster les unités de capacité de lecture sur la table DynamoDB source. L'ajout de EMR nœuds Amazon supplémentaires n'aidera pas.

Dans la sortie Hive, le pourcentage d'achèvement est mis à jour lorsqu'un ou plusieurs processus de mappers sont terminés. Pour une table DynamoDB volumineuse avec un paramètre de faible capacité allouée en lecture, la sortie du pourcentage d'achèvement peut ne pas être mise à jour pendant longtemps. Dans le cas décrit ci-dessus, la tâche apparaît complète à 0 % pendant plusieurs

heures. Pour un état plus détaillé de l'avancement de votre tâche, rendez-vous sur la EMR console Amazon ; vous pourrez consulter le statut de chaque tâche du mappeur et les statistiques relatives aux lectures de données. Vous pouvez également vous connecter à l'interface Hadoop sur le nœud maître et voir les statistiques Hadoop. Vous visualisez ainsi l'état des tâches de chaque carte et certaines statistiques de lecture des données. Pour plus d'informations, consultez les rubriques suivantes :

- [Interfaces web hébergées sur le nœud principal](#)
- [Afficher les interfaces web Hadoop](#)

Pour plus d'informations sur les exemples d'instructions HiveQL permettant d'exécuter des tâches telles que l'exportation ou l'importation de données depuis DynamoDB et la jointure de tables, consultez [Exemples de commandes Hive pour l'exportation, l'importation et l'interrogation de données dans DynamoDB](#).

Pour annuler une demande Hive

Lorsque vous exécutez une requête Hive, la réponse initiale du serveur inclut la commande pour annuler la demande. Pour annuler la demande à tout moment du processus, utilisez la Kill Command (Commande kill) de la réponse du serveur.

1. Entrez `Ctrl+C` pour quitter le client de ligne de commande.
2. À la commande shell, saisissez la Commande kill à partir de la réponse initiale du serveur à votre demande.

Vous pouvez également exécuter la commande suivante depuis la ligne de commande du nœud principal pour arrêter le job Hadoop, où *job-id* est l'identifiant de la tâche Hadoop et peut être récupéré depuis l'interface utilisateur de Hadoop.

```
hadoop job -kill job-id
```

Types de données pour Hive et DynamoDB

Le tableau suivant indique les types de données Hive disponibles, le type DynamoDB par défaut auquel ils correspondent et les autres types DynamoDB auxquels ils peuvent également mapper.

Type Hive	Type DynamoDB par défaut	Autre(s) type(s) de DynamoDB
chaîne	string (S)	
bigint ou double	number (N)	
binary	binary (B)	
boolean	booléen () BOOL	
array	list (L)	number set (NS), string set (SS) ou binary set (BS)
map<string,string>	article	map (M)
map<string,?>	map (M)	
	nul (NULL)	

Si vous voulez écrire vos données Hive sous la forme d'un autre type DynamoDB correspondant, ou si vos données DynamoDB contiennent des valeurs d'attribut d'un autre type DynamoDB, vous pouvez spécifier la colonne et le type DynamoDB avec le paramètre `dynamodb.type.mapping`. L'exemple suivant montre la syntaxe permettant de spécifier un mappage d'un autre type.

```
CREATE EXTERNAL TABLE hive_tablename (hive_column1_name column1_datatype,
hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
"hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name...",
"dynamodb.type.mapping" = "hive_column1_name:dynamodb_attribute1_datatype");
```

Le paramètre de mappage de type est facultatif et ne doit être spécifié que pour les colonnes qui utilisent d'autres types.

Par exemple, la commande Hive suivante crée une table nommée `hivetable2` qui fait référence à la table DynamoDB `dynamodbtbl2`. Elle est équivalente à `hivetable1`, sauf qu'elle mappe la colonne `col3` au type `string set (SS)`.

```
CREATE EXTERNAL TABLE hivetable2 (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtbl2",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays",
"dynamodb.type.mapping" = "col3:SS");
```

Dans Hive, `hivetable1` et `hivetable2` sont identiques. Toutefois, lorsque les données de ces tables sont écrites dans leurs tables DynamoDB correspondantes, `dynamodbtbl1` contient des listes, tandis que `dynamodbtbl2` contient des ensembles de chaînes.

Si vous souhaitez écrire des valeurs `null` Hive en tant qu'attributs de type DynamoDB pour `null`, vous pouvez le faire avec le paramètre `dynamodb.null.serialization`. L'exemple suivant montre la syntaxe permettant de spécifier la sérialisation `null`.

```
CREATE EXTERNAL TABLE hive_tablename (hive_column1_name column1_datatype,
hive_column2_name column2_datatype...)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodb_tablename",
"dynamodb.column.mapping" =
"hive_column1_name:dynamodb_attribute1_name,hive_column2_name:dynamodb_attribute2_name...",
"dynamodb.null.serialization" = "true");
```

Le paramètre de sérialisation `null` est facultatif ; il est défini sur `false` s'il n'est pas spécifié. Notez que les attributs `null` DynamoDB sont lus en tant que valeurs `null` dans Hive, quel que soit le paramètre. Les collections Hive avec des valeurs `null` peuvent être écrites dans DynamoDB uniquement si le paramètre de sérialisation `null` est spécifié comme `true`. Sinon, une erreur Hive se produit.

Le type `bigint` de Hive est identique au type `long` de Java, et le type `double` de Hive est le même que le type `double` de Java en termes de précision. Cela signifie que si vous avez des données numériques stockées dans DynamoDB qui ont une précision supérieure à celle des types de données Hive, l'utilisation de Hive pour exporter, importer ou référencer les données DynamoDB peut entraîner une perte de précision ou un échec de la requête Hive.

Les exportations de type binaire de DynamoDB vers Amazon Simple Storage Service (Amazon HDFS S3) sont également stockées sous forme de chaîne codée en Base64. Si vous importez des

données depuis Amazon S3 ou HDFS dans le type binaire DynamoDB, elles doivent être codées sous forme de chaîne Base64.

Options Hive

Vous pouvez définir les options Hive suivantes pour gérer le transfert de données hors d'Amazon DynamoDB. Ces options sont conservées uniquement pour la session Hive en cours. Si vous fermez l'invite de commande Hive et la rouvrez ultérieurement sur le cluster, ces paramètres sont rétablis à leurs valeurs par défaut.

Options Hive	Description
<code>dynamodb.throughput.read.percent</code>	<p>Définissez la vitesse des opérations de lecture pour maintenir votre débit DynamoDB dans la plage allouée pour votre table. La valeur est comprise entre 0.1 et 1.5, valeurs incluses.</p> <p>La valeur de 0,5 est la vitesse de lecture par défaut, ce qui signifie que Hive essaie de consommer la moitié des lectures provisionnées dans le cadre des ressources de la table. L'augmentation de cette valeur au-dessus de 0,5 augmente le taux de demandes de lecture. Sa diminution sous 0,5 réduit le taux de demandes de lecture. Cette vitesse de lecture est approximative. La vitesse de lecture réelle dépend de facteurs comme l'existence d'une distribution uniforme des clés dans DynamoDB.</p> <p>Si vous trouvez que votre débit alloué est fréquemment dépassé par l'opération Hive, ou si le trafic de lecture en direct est trop limité, réduisez cette valeur sous 0.5. Si vous avez suffisamment de capacité et voulez une opération Hive plus rapide, définissez cette valeur au-dessus de 0.5. Vous pouvez également sursouscrire en la définissant sur 1,5 si vous pensez qu'il existe des opérations d'entrée/sortie inutilisées disponibles.</p>

Options Hive	Description
<code>dynamodb.throughput.write.percent</code>	<p>Définissez la vitesse des opérations d'écriture pour maintenir votre débit DynamoDB dans la plage allouée pour votre table. La valeur est comprise entre 0.1 et 1.5, valeurs incluses.</p> <p>La valeur de 0,5 est la vitesse d'écriture par défaut, ce qui signifie que Hive essaie d'utiliser la moitié des écritures provisionnées dans le cadre des ressources de la table. L'augmentation de cette valeur au-dessus de 0,5 augmente le taux de demandes d'écriture. Sa diminution sous 0,5 réduit le taux de demandes d'écriture. Cette vitesse d'écriture est approximative. La vitesse d'écriture réelle dépend de facteurs comme l'existence d'une distribution uniforme des clés dans DynamoDB.</p> <p>Si vous trouvez que votre débit alloué est fréquemment dépassé par l'opération Hive ou si le trafic d'écriture en direct est trop limité, réduisez cette valeur sous 0.5. Si vous avez suffisamment de capacité et voulez une opération Hive plus rapide, définissez cette valeur au-dessus de 0.5. Vous pouvez également souscrire en la définissant jusqu'à 1,5 si vous pensez qu'il existe des opérations d'entrée/sortie inutilisées disponibles ou qu'il s'agit du chargement des données initiales sur la table et qu'il n'y pas encore de trafic en direct.</p>
<code>dynamodb.endpoint</code>	<p>Spécifiez le point de terminaison pour le service DynamoDB. Pour plus d'information sur les points de terminaison DynamoDB disponibles, consultez Régions et points de terminaison.</p>

Options Hive	Description
<code>dynamodb.max.map.tasks</code>	Spécifiez le nombre maximal de tâches de mappage lors de la lecture des données à partir de DynamoDB. Cette valeur doit être supérieure ou égale à 1.
<code>dynamodb.retry.duration</code>	Spécifiez le nombre de minutes à utiliser comme délai d'expiration pour répéter les commandes Hive. Cette valeur doit être un entier supérieur ou égal à 0. Le délai d'expiration par défaut est de deux minutes.

Ces options sont définies à l'aide de la commande SET, comme illustré dans l'exemple suivant.

```
SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

Exemples de commandes Hive pour l'exportation, l'importation et l'interrogation de données dans DynamoDB

Les exemples suivants utilisent les commandes Hive pour effectuer des opérations telles que l'exportation de données vers Amazon S3 ou HDFS l'importation de données vers DynamoDB, la jonction de tables, l'interrogation de tables, etc.

Les opérations sur une table Hive référencent les données stockées dans DynamoDB. Les commandes Hive sont soumises aux paramètres de débit alloué de la table DynamoDB et les données extraites incluent les données écrites dans la table DynamoDB au moment où la demande d'opération Hive est traitée par DynamoDB. Si le processus de récupération des données prend du temps, certaines données retournées par la commande Hive peuvent avoir été mises à jour dans DynamoDB depuis que la commande Hive a commencé.

Les commandes Hive DROP TABLE et CREATE TABLE agissent uniquement sur les tables locales de Hive et ne créent ni ne suppriment de tables dans DynamoDB. Si votre requête Hive fait référence à une table dans DynamoDB, cette table doit déjà exister avant que vous exécutiez la requête. Pour

plus d'informations sur la création et la suppression de tables dans DynamoDB, consultez [Utilisation de tables dans DynamoDB](#) dans le Guide du développeur Amazon DynamoDB.

Note

Lorsque vous mappez une table Hive avec un emplacement sur Amazon S3, ne la mappez pas avec le chemin d'accès racine du compartiment, `s3://mybucket`, car il pourrait s'ensuivre des erreurs lorsque Hive écrit les données sur Amazon S3. Au lieu de cela, mappez la table avec un sous chemin du compartiment, `s3://mybucket/mypath`.

Exportation de données à partir de DynamoDB

Vous pouvez utiliser Hive pour exporter des données à partir de DynamoDB.

Pour exporter une table DynamoDB vers un compartiment Amazon S3

- Créez une table Hive qui référence les données stockées dans DynamoDB. Vous pouvez ensuite appeler la `INSERT OVERWRITE` commande pour écrire les données dans un répertoire externe. Dans l'exemple suivant, `s3://bucketname/path/subpath/` est un chemin valide dans Amazon S3. Ajustez les colonnes et les types de données de la `CREATE` commande pour qu'ils correspondent aux valeurs de votre DynamoDB. Vous pouvez l'utiliser pour créer une archive de vos données DynamoDB dans Amazon S3.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",  
"dynamodb.column.mapping" = "col1:name, col2:year, col3:holidays");
```

```
INSERT OVERWRITE DIRECTORY 's3://bucketname/path/subpath/' SELECT *  
FROM hiveTableName;
```

Pour exporter une table DynamoDB vers un compartiment Amazon S3 à l'aide de la compartimentation.

- Créez une table externe qui fait référence à un emplacement dans Amazon S3. Cela est présenté ci-dessous en tant que `s3_export`. Au cours de l'appel `CREATE`, spécifiez le format des lignes pour le tableau. Ensuite, lorsque vous exportez des données de DynamoDB vers `s3_export`, les données sont écrites dans le format spécifié. `INSERT OVERWRITE` Dans l'exemple suivant, les données sont écrites sous forme de valeurs séparées par des virgules (`CSV`).

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodatabase1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

Pour exporter une table DynamoDB vers un compartiment Amazon S3 sans spécifier de mappage de colonne

- Créez une table Hive qui référence les données stockées dans DynamoDB. Cet exemple est similaire au précédent, sauf que vous ne spécifiez pas un mappage de colonne. La table doit avoir exactement une colonne de type `map<string, string>`. Si vous créez ensuite une table `EXTERNAL` dans Amazon S3, vous pouvez appeler la commande `INSERT OVERWRITE` pour écrire les données de DynamoDB vers Amazon S3. Vous pouvez l'utiliser pour créer une archive de vos données DynamoDB dans Amazon S3. Comme il n'y a aucun mappage de colonnes, vous ne pouvez pas interroger les tables qui sont exportées de cette façon. L'exportation de données sans spécifier de mappage de colonne est disponible dans Hive 0.8.1.5 ou version ultérieure, qui est prise en charge sur Amazon 2.2. EMR AMI x et versions ultérieures.


```

CREATE EXTERNAL TABLE hiveTableName (item map<string,string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtbl1");

CREATE EXTERNAL TABLE s3TableName (item map<string, string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3TableName SELECT *
FROM hiveTableName;

```

Pour exporter une table DynamoDB vers un compartiment Amazon S3 en utilisant la compartimentation des données

- Hive fournit plusieurs codecs de compression, que vous pouvez définir durant votre session Hive. Les données exportées sont ainsi compressées au format spécifié. L'exemple suivant compresses les fichiers exportés à l'aide de l'algorithme Lempel-Ziv-Oberhumer (). LZO

```

SET hive.exec.compress.output=true;
SET io.seqfile.compression.type=BLOCK;
SET mapred.output.compression.codec = com.hadoop.compression.lzo.LzopCodec;

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtbl1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE lzo_compression_table (line STRING)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE lzo_compression_table SELECT *
FROM hiveTableName;

```

Les codecs de compression disponibles sont les suivants :

- org.apache.hadoop.io.compress. GzipCodec
- org.apache.hadoop.io.compress. DefaultCodec
- com.hadoop.compression.lzo. LzoCodec
- com.hadoop.compression.lzo. LzopCodec
- org.apache.hadoop.io.compress. BZip2Codec
- org.apache.hadoop.io.compress. SnappyCodec

Pour exporter une table DynamoDB vers HDFS

- Utilisez la commande Hive suivante, où *hdfs:///directoryName* est un HDFS chemin valide et *hiveTableName* est une table dans Hive qui fait référence à DynamoDB. Cette opération d'exportation est plus rapide que l'exportation d'une table DynamoDB vers Amazon S3, car Hive 0.7.1.1 l'HDFS utilise comme étape intermédiaire lors de l'exportation de données vers Amazon S3. L'exemple suivant montre aussi comment définir `dynamodb.throughput.read.percent` sur 1.0 afin d'augmenter le taux de demande de lecture.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE DIRECTORY 'hdfs:///directoryName' SELECT * FROM hiveTableName;
```

Vous pouvez également exporter les données en HDFS utilisant le formatage et la compression comme indiqué ci-dessus pour l'exportation vers Amazon S3. Pour ce faire, il suffit de remplacer le répertoire Amazon S3 dans les exemples ci-dessus par un HDFS répertoire.

Pour lire des données non imprimables de UTF 8 caractères dans Hive

- Vous pouvez lire et écrire des données non UTF imprimables de 8 caractères avec Hive en utilisant la `STORED AS SEQUENCEFILE` clause lors de la création de la table. A SequenceFile

est le format de fichier binaire Hadoop ; vous devez utiliser Hadoop pour lire ce fichier. L'exemple suivant montre comment exporter des données de DynamoDB vers Amazon S3. Vous pouvez utiliser cette fonctionnalité pour gérer les caractères codés UTF -8 non imprimables.

```
CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

CREATE EXTERNAL TABLE s3_export(a_col string, b_col bigint, c_col array<string>)
STORED AS SEQUENCEFILE
LOCATION 's3://bucketname/path/subpath/';

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

Importation de données dans DynamoDB

Lorsque vous écrivez des données sur DynamoDB à l'aide de Hive, vous devez vous assurer que le nombre d'unités de capacité d'écriture est supérieur au nombre de mappeurs dans le cluster. Par exemple, les clusters qui s'exécutent sur des EC2 instances m1.xlarge produisent 8 mappeurs par instance. Dans le cas d'un cluster qui possède 10 instances, cela signifie un total de 80 mappeurs. Si vos unités de capacité en écriture ne sont pas supérieures au nombre de mappeurs dans le cluster, l'opération d'écriture Hive peut consommer la totalité du débit d'écriture ou tenter d'utiliser plus de débit qu'il n'en est alloué. Pour plus d'informations sur le nombre de mappeurs produits par chaque type d'EC2instance, consultez [Configuration de Hadoop](#).

Le nombre de mappeurs dans Hadoop est contrôlé par les divisions en entrée. S'il y a trop peu de divisions, votre commande d'écriture peut ne pas être en mesure d'utiliser tout le débit d'écriture disponible.

Si un élément avec la même clé existe dans la table DynamoDB cible, il est remplacé. Si aucun élément avec la clé n'existe dans la table DynamoDB cible, l'élément est inséré.

Pour importer une table d'Amazon S3 vers DynamoDB

- Vous pouvez utiliser Amazon EMR (AmazonEMR) et Hive pour écrire des données d'Amazon S3 vers DynamoDB.

```
CREATE EXTERNAL TABLE s3_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM s3_import;
```

Pour importer une table depuis un compartiment Amazon S3 vers DynamoDB sans spécifier de mappage de colonne

- Créez une table EXTERNAL qui référence les données stockées dans Amazon S3 et précédemment exportées depuis DynamoDB. Avant de les importer, vérifiez que la table existe dans DynamoDB et qu'elle possède le même schéma de clés que la table DynamoDB précédemment exportée. En outre, la table doit avoir une seule colonne de type map<string, string>. Si vous créez ensuite une table Hive liée à DynamoDB, vous pouvez appeler la commande INSERT OVERWRITE pour écrire les données depuis Amazon S3 vers DynamoDB. Comme il n'existe aucun mappage de colonne, vous ne pouvez pas interroger les tables qui sont importées de cette façon. L'importation de données sans spécifier de mappage de colonne est disponible dans Hive 0.8.1.5 ou version ultérieure, qui est prise en charge sur Amazon EMR AMI 2.2.3 et versions ultérieures.

```
CREATE EXTERNAL TABLE s3TableName (item map<string, string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'
LOCATION 's3://bucketname/path/subpath/';

CREATE EXTERNAL TABLE hiveTableName (item map<string, string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1");
```

```
INSERT OVERWRITE TABLE hiveTableName SELECT *
FROM s3TableName;
```

Pour importer une table depuis HDFS DynamoDB

- Vous pouvez utiliser Amazon EMR et Hive pour écrire des données depuis HDFS DynamoDB.

```
CREATE EXTERNAL TABLE hdfs_import(a_col string, b_col bigint, c_col array<string>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 'hdfs:///directoryName';

CREATE EXTERNAL TABLE hiveTableName (col1 string, col2 bigint, col3 array<string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE TABLE hiveTableName SELECT * FROM hdfs_import;
```

Interrogation de données dans DynamoDB

Les exemples suivants montrent les différentes manières dont vous pouvez utiliser Amazon EMR pour interroger les données stockées dans DynamoDB.

Pour trouver la plus grande valeur pour une colonne mappée (**max**)

- Utilisez les commandes de Hive comme suit. Dans la première commande, l'CREATE instruction crée une table Hive qui fait référence aux données stockées dans DynamoDB. L'SELECT instruction utilise ensuite cette table pour interroger les données stockées dans DynamoDB. L'exemple suivant recherche la plus grande commande passée par un client donné.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
```

```
"dynamodb.column.mapping" =  
  "customerId:CustomerId,total_cost:Cost,items_purchased:Items");  
  
SELECT max(total_cost) from hive_purchases where customerId = 717;
```

Pour regrouper les données à l'aide de la clause **GROUP BY**

- Vous pouvez utiliser la clause **GROUP BY** pour collecter les données de plusieurs enregistrements. Cette option est souvent utilisée avec une fonction d'agrégation comme somme, décompte, min ou max. L'exemple suivant retourne une liste des commandes les plus importantes des clients ayant passé plus de trois commandes.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,  
items_purchased array<String>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",  
"dynamodb.column.mapping" =  
  "customerId:CustomerId,total_cost:Cost,items_purchased:Items");  
  
SELECT customerId, max(total_cost) from hive_purchases GROUP BY customerId HAVING  
count(*) > 3;
```

Pour joindre deux tables DynamoDB

- L'exemple suivant mappe deux tables Hive avec des données stockées dans DynamoDB. Il appelle ensuite une jointure entre ces deux tables. La jointure est calculée sur le cluster, puis renvoyée. La jointure n'a pas lieu dans DynamoDB. Cet exemple renvoie une liste de clients et de leurs achats pour ceux ayant passé plus de deux commandes.

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,  
items_purchased array<String>)  
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'  
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",  
"dynamodb.column.mapping" =  
  "customerId:CustomerId,total_cost:Cost,items_purchased:Items");
```

```
CREATE EXTERNAL TABLE hive_customers(customerId bigint, customerName string,
customerAddress array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Customers",
"dynamodb.column.mapping" =
"customerId:CustomerId,customerName:Name,customerAddress:Address");

Select c.customerId, c.customerName, count(*) as count from hive_customers c
JOIN hive_purchases p ON c.customerId=p.customerId
GROUP BY c.customerId, c.customerName HAVING count > 2;
```

Pour joindre deux tables de différentes sources

- Dans l'exemple suivant, *Customer_S3* est une table Hive qui charge un CSV fichier stocké dans Amazon S3 et *hive_purchases* est une table qui référence des données dans DynamoDB. L'exemple suivant réunit les données clients stockées sous forme de CSV fichier dans Amazon S3 avec les données de commande stockées dans DynamoDB pour renvoyer un ensemble de données représentant les commandes passées par des clients dont le nom contient « Miller ».

```
CREATE EXTERNAL TABLE hive_purchases(customerId bigint, total_cost double,
items_purchased array<String>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "Purchases",
"dynamodb.column.mapping" =
"customerId:CustomerId,total_cost:Cost,items_purchased:Items");

CREATE EXTERNAL TABLE Customer_S3(customerId bigint, customerName string,
customerAddress array<String>)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION 's3://bucketname/path/subpath/';

Select c.customerId, c.customerName, c.customerAddress from
Customer_S3 c
JOIN hive_purchases p
ON c.customerid=p.customerid
where c.customerName like '%Miller%';
```

Note

Dans les exemples précédents, les CREATE TABLE déclarations ont été incluses dans chaque exemple par souci de clarté et d'exhaustivité. Lorsque vous exécutez plusieurs requêtes ou opérations d'exportation sur une table Hive donnée, vous ne devez créer la table qu'une seule fois, au début de la session Hive.

Optimisation des performances pour les EMR opérations Amazon dans DynamoDB

EMR Les opérations Amazon sur une table DynamoDB sont considérées comme des opérations de lecture et sont soumises aux paramètres de débit provisionnés de la table. Amazon EMR met en œuvre sa propre logique pour essayer d'équilibrer la charge sur votre table DynamoDB afin de minimiser le risque de dépassement du débit alloué. À la fin de chaque requête Hive, Amazon EMR renvoie des informations sur le cluster utilisé pour traiter la requête, notamment le nombre de fois où le débit alloué a été dépassé. Vous pouvez utiliser ces informations, ainsi que CloudWatch les mesures relatives à votre débit DynamoDB, pour mieux gérer la charge sur votre table DynamoDB lors des demandes suivantes.

Les facteurs suivants influencent les performances des requêtes Hive lorsque vous travaillez avec des tables DynamoDB.

Unités de capacité de lecture allouées

Lorsque vous exécutez les requêtes Hive sur une table DynamoDB, vous devez vous assurer que vous avez alloué une quantité suffisante d'unités de capacité de lecture.

Par exemple, supposons que vous ayez provisionné 100 unités de capacité en lecture pour votre table DynamoDB. Vous pourrez ainsi effectuer 100 lectures, ou 409 600 octets, par seconde. Si cette table contient 20 Go de données (21 474 836 480 octets) et que votre requête Hive exécute une analyse complète de la table, vous pouvez estimer la durée d'exécution de la requête :

$$21\,474\,836\,480 / 409\,600 = 52\,429 \text{ secondes} = 14,56 \text{ heures}$$

Le seul moyen de diminuer le temps nécessaire consiste à ajuster les unités de capacité de lecture sur la table DynamoDB source. L'ajout de nœuds supplémentaires au EMR cluster Amazon n'aidera pas.

Dans la sortie Hive, le pourcentage d'achèvement est mis à jour lorsqu'un ou plusieurs processus de mappers sont terminés. Pour une table DynamoDB volumineuse avec un paramètre de faible capacité allouée en lecture, la sortie du pourcentage d'achèvement peut ne pas être mise à jour pendant longtemps. Dans le cas décrit ci-dessus, la tâche apparaît complète à 0 % pendant plusieurs heures. Pour un état plus détaillé de l'avancement de votre tâche, rendez-vous sur la EMR console Amazon ; vous pourrez consulter le statut de chaque tâche du mapper et les statistiques relatives aux lectures de données.

Vous pouvez également vous connecter à l'interface Hadoop sur le nœud maître et voir les statistiques Hadoop. Vous visualisez ainsi l'état des tâches de chaque carte et certaines statistiques de lecture des données. Pour plus d'informations, consultez la section [Interfaces Web hébergées sur le nœud principal](#) dans l'Amazon EMR Management Guide.

Paramètre de pourcentage de lecture

Par défaut, Amazon EMR gère la charge des demandes par rapport à votre table DynamoDB en fonction de votre débit provisionné actuel. Toutefois, lorsqu'Amazon EMR renvoie des informations relatives à votre tâche qui incluent un nombre élevé de réponses ayant dépassé le débit alloué, vous pouvez ajuster le taux de lecture par défaut à l'aide du `dynamodb.throughput.read.percent` paramètre lorsque vous configurez la table Hive. Pour plus d'informations sur la configuration du paramètre de pourcentage de lecture, consultez [Options Hive](#).

Paramètre de pourcentage d'écriture

Par défaut, Amazon EMR gère la charge des demandes par rapport à votre table DynamoDB en fonction de votre débit provisionné actuel. Toutefois, lorsqu'Amazon EMR renvoie des informations relatives à votre tâche qui incluent un nombre élevé de réponses ayant dépassé le débit alloué, vous pouvez ajuster le taux d'écriture par défaut à l'aide du `dynamodb.throughput.write.percent` paramètre lorsque vous configurez la table Hive. Pour plus d'informations sur la configuration du paramètre de pourcentage d'écriture, consultez [Options Hive](#).

Paramètre de durée de nouvelle tentative

Par défaut, Amazon EMR réexécute une requête Hive si elle n'a pas renvoyé de résultat dans les deux minutes, ce qui correspond à l'intervalle de tentatives par défaut. Vous pouvez ajuster cet intervalle en définissant le paramètre `dynamodb.retry.duration` lorsque vous exécutez une requête Hive. Pour plus d'informations sur la configuration du paramètre de pourcentage d'écriture, consultez [Options Hive](#).

Nombre de tâches de mappage

Les démons de mappeur que Hadoop lance pour traiter vos requêtes afin d'exporter et d'interroger des données stockées dans DynamoDB sont limités à une vitesse de lecture maximale de 1 Mio par seconde afin de limiter la capacité de lecture utilisée. Si vous avez un débit alloué supplémentaire disponible sur DynamoDB, vous pouvez améliorer les performances de l'exportation Hive et des opérations d'interrogation en augmentant le nombre de démons de mappeur. Pour ce faire, vous pouvez soit augmenter le nombre d'EC2instances de votre cluster, soit augmenter le nombre de démons de mappage exécutés sur chaque instance. EC2

Vous pouvez augmenter le nombre d'EC2instances d'un cluster en arrêtant le cluster actuel et en le relançant avec un plus grand nombre d'EC2instances. Vous spécifiez le nombre d'EC2instances dans la boîte de dialogue Configurer les EC2 instances si vous lancez le cluster depuis la EMR console Amazon, ou avec l'`--num-instances` option si vous lancez le cluster depuis le CLI.

Le nombre de tâches cartographiques exécutées sur une instance dépend du type d'EC2instance. Pour plus d'informations sur les types d'EC2instances pris en charge et le nombre de mappeurs fournis par chacun d'entre eux, consultez [Configuration de la tâche](#). Là, vous trouverez une section « Configuration des tâches » pour chacune des configurations prises en charge.

Un autre moyen d'augmenter le nombre de démons de mappeur consiste à modifier le paramètre de configuration `mapreduce.tasktracker.map.tasks.maximum` de Hadoop sur une valeur plus élevée. Cela a l'avantage de vous donner plus de mappeurs sans augmenter le nombre ou la taille des EC2 instances, ce qui vous permet d'économiser de l'argent. L'inconvénient est que si cette valeur est trop élevée, les EC2 instances de votre cluster risquent de manquer de mémoire. Pour définir `mapreduce.tasktracker.map.tasks.maximum`, lancez le cluster et spécifiez une valeur pour `mapreduce.tasktracker.map.tasks.maximum` en tant que propriété de la classification de configuration `mapred-site`. Voici un exemple : Pour de plus amples informations, veuillez consulter [Configuration des applications](#).

```
{
  "configurations": [
    {
      "classification": "mapred-site",
      "properties": {
        "mapred.tasktracker.map.tasks.maximum": "10"
      }
    }
  ]
}
```

}

Demandes de données en parallèle

Plusieurs demandes de données, à partir de plusieurs utilisateurs ou de plusieurs applications vers une seule table, peuvent diminuer la vitesse de lecture allouée et ralentir les performances.

Durée du processus

La cohérence des données dans DynamoDB dépend de l'ordre des opérations de lecture et d'écriture sur chaque nœud. Quand une requête Hive est en cours, une autre application peut charger de nouvelles données dans la table DynamoDB, voire modifier ou supprimer des données existantes. Dans ce cas, les résultats de la requête Hive peuvent ne pas tenir compte des modifications apportées aux données pendant l'exécution de la requête.

Eviter le dépassement de débit

Lors de l'exécution de requêtes Hive sur DynamoDB, veillez à ne pas dépasser votre débit alloué, vous risquez sinon de réduire la capacité nécessaire des appels à DynamoDB : :Get de votre application. Pour éviter que cela ne se produise, vous devez surveiller régulièrement le volume de lecture et la limitation des appels aux applications DynamoDB : :Get en consultant les journaux et les métriques de surveillance sur Amazon. CloudWatch

Durée de la demande

La planification des requêtes Hive qui accèdent à une table DynamoDB à un moment où la demande sur celle-ci est inférieure a pour effet d'améliorer les performances. Par exemple, si la plupart des utilisateurs de votre application vivent à San Francisco, vous pouvez choisir d'exporter les données chaque jour à 4h00 PST, lorsque la majorité des utilisateurs sont endormis et ne mettent pas à jour les enregistrements de votre base de données DynamoDB.

Tables basées sur le temps

Si les données sont organisées en tant que série de tables DynamoDB basées sur le temps, telles qu'une table par jour, vous pouvez exporter les données lorsque la table n'est plus active. Vous pouvez utiliser cette technique pour sauvegarder les données sur Amazon S3 en continu.

Données archivées

Si vous prévoyez d'exécuter de nombreuses requêtes Hive sur les données stockées dans DynamoDB et que votre application peut tolérer les données archivées, vous souhaitez peut-

être exporter les données vers HDFS Amazon S3 et exécuter les requêtes Hive sur une copie des données plutôt que sur DynamoDB. Cela permet de conserver vos opérations de lecture et votre débit alloué.

Kinesis

EMR Les clusters Amazon peuvent lire et traiter les flux Amazon Kinesis directement à l'aide d'outils familiers de l'écosystème Hadoop tels que Hive, Pig MapReduce, Hadoop Streaming et Cascading. API Vous pouvez également associer les données en temps réel d'Amazon Kinesis aux données existantes sur Amazon S3, Amazon DynamoDB et dans un cluster en cours d'exécution. HDFS Vous pouvez charger directement les données d'Amazon EMR vers Amazon S3 ou DynamoDB pour les activités de post-traitement. Pour plus d'informations sur les caractéristiques principales du service Amazon Kinesis et ses tarifs, voir la page [Amazon Kinesis](#).

Que puis-je faire avec l'intégration entre Amazon EMR et Amazon Kinesis ?

L'intégration entre Amazon EMR et Amazon Kinesis facilite considérablement certains scénarios, par exemple :

- Analyse de journaux de streaming – Vous pouvez analyser les journaux web de vos flux de diffusion continus afin de générer la liste des 10 principaux types d'erreurs par région, par navigateur et par domaine d'accès à intervalles de quelques minutes.
- Engagement client – Vous pouvez écrire des requêtes associant des données de parcours provenant d'Amazon Kinesis à des informations sur une campagne publicitaire stockées dans une table DynamoDB, dans le but d'identifier les catégories de publicité les plus efficaces parmi celles affichées sur des sites web donnés.
- Requêtes interactives ad hoc : vous pouvez charger régulièrement des données provenant de flux HDFS Amazon Kinesis et les rendre disponibles sous forme de table Impala locale pour des requêtes analytiques rapides et interactives.

Analyse de points de contrôle des flux Amazon Kinesis

Les utilisateurs peuvent exécuter des analyses régulières par lots de flux Amazon Kinesis dans ce que l'on appelle des itérations. Comme les enregistrements de données de flux Amazon Kinesis sont extraits à l'aide d'un numéro de séquence, les limites d'itération sont définies par des numéros de séquence de début et de fin qu'Amazon EMR stocke dans une table DynamoDB. Par exemple, quand

`iteration0` se termine, elle stocke le numéro de fin de séquence dans DynamoDB afin que lorsque la tâche `iteration1` commence, elle puisse extraire les données suivantes du flux. Ce mappage d'itérations de données de flux s'appelle un point de contrôle. Pour plus d'informations, consultez [Connecteur Kinesis](#).

Si une itération a été vérifiée et que la tâche a échoué lors du traitement d'une itération, Amazon EMR tente de retraiter les enregistrements de cette itération.

Les points de contrôle sont une fonctionnalité qui vous permet de :

- Commencer le traitement de données après un numéro de séquence traité par une requête précédente qui s'est exécutée sur le même flux de données et nom logique
- Retraiter le même lot de données à partir de Kinesis qui a été traité par une requête précédente

Pour activer les points de contrôle, définissez le paramètre `kinesis.checkpoint.enabled` sur `true` dans vos scripts. En outre, configurez les paramètres suivants :

Paramètre de configuration	Description
<code>kinesis.checkpoint.metastore.table.name</code>	Nom de la table DynamoDB où seront stockées les informations de point de contrôle
<code>kinesis.checkpoint.metastore.hash.key.name</code>	Nom de la clé de hachage pour la table DynamoDB
<code>kinesis.checkpoint.metastore.hash.range.name</code>	Nom de la clé de plage pour la table DynamoDB
<code>kinesis.checkpoint.logical.name</code>	Un nom logique pour le traitement actuel
<code>kinesis.checkpoint.iteration.no</code>	Numéro de l'itération pour le traitement associé au nom logique
<code>kinesis.rerun.iteration.without.wait</code>	Valeur booléenne qui indique si une itération ayant échoué peut être exécutée à nouveau sans attendre l'expiration ; la valeur par défaut est <code>false</code>

IOPSRecommandations provisionnées pour les tables Amazon DynamoDB

Le EMR connecteur Amazon pour Amazon Kinesis utilise la base de données DynamoDB comme support pour les métadonnées de point de contrôle. Vous devez créer une table dans DynamoDB avant de consommer des données dans un flux Amazon Kinesis avec un EMR cluster Amazon à intervalles réguliers. La table doit se trouver dans la même région que votre EMR cluster Amazon. Voici des recommandations générales concernant le nombre de tâches IOPS que vous devez prévoir pour vos tables DynamoDB ; j soit le nombre maximum de tâches Hadoop (avec une combinaison nom logique+numéro d'itération différente) pouvant être exécutées simultanément s et le nombre maximal de partitions que chaque tâche peut traiter :

Pour Read Capacity Units (Unités de capacité en lecture) : $j * s / 5$

Pour Write Capacity Units (Unités de capacité en écriture) : $j * s$

Considérations sur les performances

Le débit des partitions Amazon Kinesis est directement proportionnel à la taille de l'instance des nœuds dans les EMR clusters Amazon et à la taille des enregistrements dans le flux. Nous vous recommandons d'utiliser des instances m5.xlarge ou plus importantes sur les nœuds maîtres et principaux.

Planifiez une analyse Amazon Kinesis avec Amazon EMR

Lorsque vous analysez des données sur un flux Amazon Kinesis actif, limité par des délais d'attente et une durée maximale pour l'itération, il est important que vous exécutiez l'analyse fréquemment pour collecter régulièrement des détails à partir du flux. Il existe plusieurs façons d'exécuter ces scripts et requêtes à intervalles réguliers ; nous vous conseillons d'utiliser [AWS Data Pipeline PigActivity](#) et consultez [AWS Data Pipeline HiveActivity](#) le Guide du AWS Data Pipeline développeur.

Migration du connecteur Spark Kinesis vers la version 2.x SDK pour Amazon 7.0 EMR

AWS SDKII fournit un ensemble complet de bibliothèques permettant APIs d'interagir avec les services de AWS cloud computing, tels que la gestion des informations d'identification, la connexion aux services S3 et Kinesis. Le connecteur Spark Kinesis est utilisé pour consommer les données des flux de données Kinesis, et les données reçues sont transformées et traitées dans le moteur

d'exécution de Spark. Actuellement, ce connecteur est construit sur 1.x de AWS SDK et Kinesis-client-library (KCL).

Dans le cadre de la migration vers la version AWS SDK 2.x, le connecteur Spark Kinesis est également mis à jour en conséquence pour fonctionner avec SDK la version 2.x. Dans la version Amazon EMR 7.0, Spark contient la mise à niveau SDK 2.x qui n'est pas encore disponible dans la version communautaire d'Apache Spark. Si vous utilisez le connecteur Spark Kinesis à partir d'une version inférieure à 7.0, vous devez migrer les codes de vos applications pour qu'ils s'exécutent sur SDK 2.x avant de pouvoir migrer vers Amazon 7.0. EMR

Guides de migration

Cette section décrit les étapes de migration d'une application vers la version mise à niveau du connecteur Spark Kinesis. Il inclut des guides de migration vers la Kinesis Client Library (KCL) 2.x, des fournisseurs AWS d'identifiants et des clients de AWS service dans la version 2.x. AWS SDK À titre de référence, il inclut également un exemple de [WordCount](#) programme utilisant le connecteur Kinesis.

Rubriques

- [Migration de la version 1.x KCL vers la version 2.x](#)
- [Migration des fournisseurs d' AWS informations d'identification de la version AWS SDK 1.x vers la version 2.x](#)
- [Migration des clients de AWS service de la version AWS SDK 1.x à la version 2.x](#)
- [Exemples de code pour les applications de streaming](#)
- [Considérations relatives à l'utilisation de la version mise à niveau du connecteur Spark Kinesis](#)

Migration de la version 1.x KCL vers la version 2.x

- Niveau et dimensions des métriques dans **KinesisInputDStream**

Lorsque vous instanciez un `KinesisInputDStream`, vous pouvez contrôler le niveau et les dimensions des métriques du flux. L'exemple suivant montre comment personnaliser ces paramètres avec la version KCL 1.x :

```
import
  com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration
import com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel
```

```

val kinesisStream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(kinesisCheckpointInterval)
  .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
  .metricsLevel(MetricsLevel.DETAILED)

  .metricsEnabledDimensions(KinesisClientLibConfiguration.DEFAULT_METRICS_ENABLED_DIMENSIONS.a
  .build()

```

Dans la KCL version 2.x, ces paramètres de configuration ont des noms de package différents. Pour migrer vers la version 2.x :

1. Remplacez les instructions d'importation pour `com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration` et `com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel` par `software.amazon.kinesis.metrics.MetricsLevel` et `software.amazon.kinesis.metrics.MetricsUtil` respectivement.

```

// import com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel
import software.amazon.kinesis.metrics.MetricsLevel

// import
  com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibConfiguration
import software.amazon.kinesis.metrics.MetricsUtil

```

2. Remplacez la ligne `metricsEnabledDimensionsKinesisClientLibConfiguration.DEFAULT_METRICS_ENABLED_DIMENSIONS` par `metricsEnabledDimensionsSet(MetricsUtil.OPERATION_DIMENSION_NAME, MetricsUtil.SHARD_ID_DIMENSION_NAME)`.

Vous trouverez ci-dessous une version mise à jour du `KinesisInputDStream` avec un niveau de métrique et des dimensions de métriques personnalisés :

```

import software.amazon.kinesis.metrics.MetricsLevel
import software.amazon.kinesis.metrics.MetricsUtil

```



```
val kinesisStream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(kinesisCheckpointInterval)
  .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
  .metricsLevel(MetricsLevel.DETAILED)
  .metricsEnabledDimensions(Set(MetricsUtil.OPERATION_DIMENSION_NAME,
    MetricsUtil.SHARD_ID_DIMENSION_NAME))
  .build()
```

- Fonction de gestionnaire de messages dans KinesisInputDStream

Lors de l'instanciation d'un `KinesisInputDStream`, vous pouvez également spécifier une « fonction de gestionnaire de messages » qui prend un enregistrement Kinesis et renvoie un objet générique `T`, au cas où vous souhaiteriez utiliser d'autres données incluses dans un enregistrement, comme une clé de partition.

Dans la KCL version 1.x, la signature de la fonction du gestionnaire de messages est `:Record => T`, où se trouve `Record`. `com.amazonaws.services.kinesis.model.Record`

Dans la KCL version 2.x, la signature du gestionnaire est remplacée par `:KinesisClientRecord => T`, where `KinesisClientRecord` is. `software.amazon.kinesis.retrieval.KinesisClientRecord`

Voici un exemple de fourniture d'un gestionnaire de messages dans la version KCL 1.x :

```
import com.amazonaws.services.kinesis.model.Record

def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
val stream = KinesisInputDStream.builder
  .streamingContext(ssc)
  .streamName(streamName)
  .endpointUrl(endpointUrl)
  .regionName(regionName)
  .initialPosition(new Latest())
  .checkpointAppName(appName)
  .checkpointInterval(Seconds(10))
```

```
.storageLevel(StorageLevel.MEMORY_ONLY)
.buildWithMessageHandler(addFive)
```

Pour migrer le gestionnaire de messages :

1. Remplacez l'instruction d'importation `com.amazonaws.services.kinesis.model.Record` par `software.amazon.kinesis.retrieval.KinesisClientRecord`.

```
// import com.amazonaws.services.kinesis.model.Record
import software.amazon.kinesis.retrieval.KinesisClientRecord
```

2. Mettez à jour la signature de la méthode du gestionnaire de messages.

```
//def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt
+ 5
```

Voici un exemple mis à jour de fourniture du gestionnaire de messages dans la version KCL 2.x :

```
import software.amazon.kinesis.retrieval.KinesisClientRecord

def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt + 5
val stream = KinesisInputDStream.builder
    .streamingContext(ssc)
    .streamName(streamName)
    .endpointUrl(endpointUrl)
    .regionName(regionName)
    .initialPosition(new Latest())
    .checkpointAppName(appName)
    .checkpointInterval(Seconds(10))
    .storageLevel(StorageLevel.MEMORY_ONLY)
    .buildWithMessageHandler(addFive)
```

Pour plus d'informations sur la migration de la version KCL 1.x vers la version 2.x, consultez la section [Migration des consommateurs de la version 1.x vers KCL la version 2.x](#). KCL

Migration des fournisseurs d' AWS informations d'identification de la version AWS SDK 1.x vers la version 2.x

Les fournisseurs d'informations d'identification sont utilisés pour obtenir des AWS informations d'identification pour les interactions avec AWS. Il existe plusieurs modifications d'interface et de classe liées aux fournisseurs d'informations d'identification dans la SDK version 2.x, qui peuvent être consultées [ici](#). Le connecteur Spark Kinesis a défini une interface (`org.apache.spark.streaming.kinesis.SparkAWSCredentials`) et des classes d'implémentation qui renvoient la version 1.x des fournisseurs d'informations d' AWS identification. Ces fournisseurs d'informations d'identification sont nécessaires lors de l'initialisation des clients Kinesis. Par exemple, si vous utilisez cette méthode `SparkAWSCredentials.provider` dans les applications, vous devrez mettre à jour les codes pour utiliser la version 2.x des fournisseurs AWS d'informations d'identification.

Voici un exemple d'utilisation des fournisseurs d'informations d'identification dans la version AWS SDK 1.x :

```
import org.apache.spark.streaming.kinesis.SparkAWSCredentials
import com.amazonaws.auth.AWSCredentialsProvider

val basicSparkCredentials = SparkAWSCredentials.builder
    .basicCredentials("accessKey", "secretKey")
    .build()

val credentialProvider = basicSparkCredentials.provider
assert(credentialProvider.isInstanceOf[AWSCredentialsProvider], "Type should be
    AWSCredentialsProvider")
```

Pour migrer vers la version SDK 2.x :

1. Remplacez l'instruction d'importation `com.amazonaws.auth.AWSCredentialsProvider` par `software.amazon.awssdk.auth.credentials.AwsCredentialsProvider`.

```
//import com.amazonaws.auth.AWSCredentialsProvider
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider
```

2. Mettez à jour les autres codes qui utilisent cette classe.

```
import org.apache.spark.streaming.kinesis.SparkAWSCredentials
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider
```

```
val basicSparkCredentials = SparkAWSCredentials.builder
    .basicCredentials("accessKey", "secretKey")
    .build()

val credentialProvider = basicSparkCredentials.provider
assert (credentialProvider.isInstanceOf[AwsCredentialsProvider], "Type should be
    AwsCredentialsProvider")
```

Migration des clients de AWS service de la version AWS SDK 1.x à la version 2.x

AWS les clients de service ont des noms de package différents dans la version 2.x (c'est-à-dire `software.amazon.awssdk`), alors que dans la version SDK 1.x les utilise `com.amazonaws`. Pour plus d'informations sur les modifications apportées aux clients, voir [cette page](#). Si vous utilisez ces clients de service dans les codes, vous devez migrer les clients en conséquence.

Voici un exemple de création d'un client dans la version SDK 1.x :

```
import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
import com.amazonaws.services.dynamodbv2.document.DynamoDB

AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();
```

Pour migrer vers la version 2.x :

1. Modifiez les instructions d'importation pour les clients de service. Prenons l'exemple des clients DynamoDB. Remplacez `com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient` ou `com.amazonaws.services.dynamodbv2.document.DynamoDB` par `software.amazon.awssdk.services.dynamodb.DynamoDbClient`.

```
// import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
// import com.amazonaws.services.dynamodbv2.document.DynamoDB
import software.amazon.awssdk.services.dynamodb.DynamoDbClient
```

2. Mettez à jour les codes qui initialisent les clients.

```
// AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
// AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();
```

```
DynamoDbClient ddbClient = DynamoDbClient.create();
DynamoDbClient ddbClient = DynamoDbClient.builder().build();
```

Pour plus d'informations sur la migration AWS SDK de la version 1.x vers la version 2.x, consultez la section [Quelles sont les différences entre les versions 1.x et 2.x AWS SDK pour Java ?](#)

Exemples de code pour les applications de streaming

```
import java.net.URI
import software.amazon.awssdk.auth.credentials.DefaultCredentialsProvider
import software.amazon.awssdk.http.apache.ApacheHttpClient
import software.amazon.awssdk.services.kinesis.KinesisClient
import software.amazon.awssdk.services.kinesis.model.DescribeStreamRequest
import software.amazon.awssdk.regions.Region
import software.amazon.kinesis.metrics.{MetricsLevel, MetricsUtil}

import org.apache.spark.SparkConf
import org.apache.spark.storage.StorageLevel
import org.apache.spark.streaming.{Milliseconds, StreamingContext}
import org.apache.spark.streaming.dstream.DStream.toPairDStreamFunctions
import org.apache.spark.streaming.kinesis.KinesisInitialPositions.Latest
import org.apache.spark.streaming.kinesis.KinesisInputDStream

object KinesisWordCountASLSDKV2 {

  def main(args: Array[String]): Unit = {
    val appName = "demo-app"
    val streamName = "demo-kinesis-test"
    val endpointUrl = "https://kinesis.us-west-2.amazonaws.com"
    val regionName = "us-west-2"

    // Determine the number of shards from the stream using the low-level Kinesis
    Client
    // from the AWS Java SDK.
    val credentialsProvider = DefaultCredentialsProvider.create
    require(credentialsProvider.resolveCredentials() != null,
      "No AWS credentials found. Please specify credentials using one of the methods
    specified " +
```

```
    "in https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/
credentials.html")
    val kinesisClient = KinesisClient.builder()
        .credentialsProvider(credentialsProvider)
        .region(Region.US_WEST_2)
        .endpointOverride(URI.create(endpointUrl))
        .httpClientBuilder(ApacheHttpClient.builder())
        .build()
    val describeStreamRequest = DescribeStreamRequest.builder()
        .streamName(streamName)
        .build()
    val numShards = kinesisClient.describeStream(describeStreamRequest)
        .streamDescription
        .shards
        .size

    // In this example, we are going to create 1 Kinesis Receiver/input DStream for
    each shard.
    // This is not a necessity; if there are less receivers/DStreams than the number of
    shards,
    // then the shards will be automatically distributed among the receivers and each
    receiver
    // will receive data from multiple shards.
    val numStreams = numShards

    // Spark Streaming batch interval
    val batchInterval = Milliseconds(2000)

    // Kinesis checkpoint interval is the interval at which the DynamoDB is updated
    with information
    // on sequence number of records that have been received. Same as batchInterval for
    this
    // example.
    val kinesisCheckpointInterval = batchInterval

    // Setup the SparkConfig and StreamingContext
    val sparkConfig = new SparkConf().setAppName("KinesisWordCountASLSDKV2")
    val ssc = new StreamingContext(sparkConfig, batchInterval)

    // Create the Kinesis DStreams
    val kinesisStreams = (0 until numStreams).map { i =>
        KinesisInputDStream.builder
            .streamingContext(ssc)
```

```

        .streamName(streamName)
        .endpointUrl(endpointUrl)
        .regionName(regionName)
        .initialPosition(new Latest())
        .checkpointAppName(appName)
        .checkpointInterval(kinesisCheckpointInterval)
        .storageLevel(StorageLevel.MEMORY_AND_DISK_2)
        .metricsLevel(MetricsLevel.DETAILED)
        .metricsEnabledDimensions(Set(MetricsUtil.OPERATION_DIMENSION_NAME,
MetricsUtil.SHARD_ID_DIMENSION_NAME))
        .build()
    }

    // Union all the streams
    val unionStreams = ssc.union(kinesisStreams)

    // Convert each line of Array[Byte] to String, and split into words
    val words = unionStreams.flatMap(byteArray => new String(byteArray).split(" "))

    // Map each word to a (word, 1) tuple so we can reduce by key to count the words
    val wordCounts = words.map(word => (word, 1)).reduceByKey(_ + _)

    // Print the first 10 wordCounts
    wordCounts.print()

    // Start the streaming context and await termination
    ssc.start()
    ssc.awaitTermination()
}
}

```

Considérations relatives à l'utilisation de la version mise à niveau du connecteur Spark Kinesis

- Si vos applications utilisent la JDK version `Kinesis-producer-library` avec une version inférieure à 11, vous pouvez rencontrer des exceptions telles que `java.lang.NoClassDefFoundError: javax/xml/bind/DatatypeConverter`. Cela se produit parce que la EMR version 7.0 est fournie avec JDK 17 par défaut et que les modules J2EE ont été supprimés des bibliothèques standard depuis Java 11+. Il est possible de résoudre le problème en ajoutant la dépendance suivante dans le fichier pom. Remplacez la version de la bibliothèque par une autre selon vos besoins.

```
<dependency>
```

```
<groupId>javax.xml.bind</groupId>
<artifactId>jaxb-api</artifactId>
<version>${jaxb-api.version}</version>
</dependency>
```

- Le fichier jar du connecteur Spark Kinesis se trouve sous ce chemin après la création d'un EMR cluster : `/usr/lib/spark/connector/lib/`

S3 DistCp (s3-dist-cp)

Apache DistCp est un outil open source que vous pouvez utiliser pour copier de grandes quantités de données. S3 DistCp est similaire DistCp, mais optimisé pour fonctionner avec AWS, en particulier, Amazon S3. La commande pour S3 DistCp dans Amazon EMR version 4.0 et versions ultérieures est `s3-dist-cp`, que vous ajoutez en tant qu'étape dans un cluster ou sur la ligne de commande. Grâce à S3DistCp, vous pouvez copier efficacement de grandes quantités de données d'Amazon S3 vers HDFS lesquelles elles peuvent être traitées par les étapes suivantes dans votre EMR cluster Amazon. Vous pouvez également utiliser S3 DistCp pour copier des données entre des compartiments Amazon S3 ou depuis HDFS Amazon S3. S3 DistCp est plus évolutif et plus efficace pour copier en parallèle un grand nombre d'objets entre des compartiments et des AWS comptes.

Pour des commandes spécifiques démontrant la flexibilité de S3DistCP dans des scénarios réels, consultez [Sept conseils d'utilisation de S3 DistCp](#) sur le AWS blog Big Data.

Par exemple DistCp, S3 DistCp utilise MapReduce pour copier de manière distribuée. Il partage les tâches de copie, de gestion des erreurs, de récupération et de création de rapports entre plusieurs serveurs. Pour plus d'informations sur le projet DistCp open source Apache, consultez le [DistCp guide](#) dans la documentation d'Apache Hadoop.

Si S3 DistCp ne parvient pas à copier tout ou partie des fichiers spécifiés, l'étape du cluster échoue et renvoie un code d'erreur différent de zéro. Dans ce cas, S3 DistCp ne nettoie pas les fichiers partiellement copiés.

Important

S3 DistCp ne prend pas en charge les noms de compartiment Amazon S3 contenant le caractère de soulignement.

S3 DistCp ne prend pas en charge la concaténation pour les fichiers Parquet. Utilisez PySpark plutôt. Pour plus d'informations, consultez la section [Concaténation de fichiers de parquet](#) sur Amazon. EMR

Pour éviter les erreurs de copie lorsque vous utilisez S3DistCP pour copier un seul fichier (au lieu d'un répertoire) depuis S3 vers, HDFS utilisez Amazon EMR version 5.33.0 ou ultérieure, ou Amazon EMR version 6.3.0 ou ultérieure.

DistCp Options S3

Bien que similaire à DistCp, S3 DistCp prend en charge un ensemble d'options différent pour modifier la façon dont il copie et compresse les données.

Lorsque vous appelez S3DistCp, vous pouvez spécifier les options décrites dans le tableau suivant. Les options sont ajoutées à l'étape à l'aide de la liste d'arguments. Des exemples d'arguments S3 DistCp sont présentés dans le tableau suivant.

Option	Description	Obligatoire
<code>--src=LOCATION</code>	<p>Emplacement des données à copier. Il peut s'agir d'un emplacement HDFS ou d'un emplacement Amazon S3.</p> <p>Exemple : <code>--src=s3://amzn-s3-demo-bucket1/logs/j-3GYXXXXXX9I0J/node</code></p> <div style="border: 1px solid #f08080; border-radius: 10px; padding: 10px; margin-top: 10px;"> <p>⚠ Important</p> <p>S3 DistCp ne prend pas en charge les noms de compartiment Amazon S3 contenant le caractère de soulignement.</p> </div>	Oui
<code>--dest=LOCATION</code>	<p>Destination des données. Il peut s'agir d'un emplacement HDFS ou d'un emplacement Amazon S3.</p> <p>Exemple : <code>--dest=hdfs:///output</code></p>	Oui

Option	Description	Obligatoire
	<div style="border: 1px solid #f08080; padding: 10px; background-color: #fff9e6;"> <p>⚠ Important</p> <p>S3 DistCp ne prend pas en charge les noms de compartiment Amazon S3 contenant le caractère de soulignement.</p> </div>	
<pre>--srcPattern=PATTERN</pre>	<p><u>Expression régulière</u> qui filtre l'opération de copie sur un sous-ensemble des données à l'adresse <code>--src</code>. Si <code>--srcPattern</code> et <code>--groupBy</code> ne sont pas spécifiés, toutes les données à l'adresse <code>--src</code> sont copiées dans <code>--dest</code>.</p> <p>Si l'argument de l'expression régulière contient des caractères spéciaux, tels qu'un astérisque (*), l'expression régulière ou la chaîne <code>--args</code> complète doit être mise entre guillemets simples (').</p> <p>Exemple : <code>--srcPattern=.*daemons.*-ha doop-.*</code></p>	Non

Option	Description	Obligatoire
<pre>--groupBy=PATTERN</pre>	<p>Expression régulière qui oblige S3 DistCp à concaténer les fichiers correspondant à l'expression. Par exemple, vous pouvez utiliser cette option pour combiner tous les fichiers journaux écrits en une heure dans un fichier unique. Le nom de fichier concaténé est la valeur correspondante à l'expression régulière pour le regroupement.</p> <p>Les parenthèses indiquent comment les fichiers doivent être regroupés, en combinant tous les éléments qui correspondent à l'instruction entre parenthèses en un fichier de sortie unique. Si l'expression régulière n'inclut pas d'instruction entre parenthèses, le cluster échoue à l'DistCp étape S3 et renvoie une erreur.</p> <p>Si l'argument de l'expression régulière contient des caractères spéciaux, tels qu'un astérisque (*), l'expression régulière ou la chaîne <code>--args</code> complète doit être mise entre guillemets simples (').</p> <p>Lorsque <code>--groupBy</code> est spécifié, seuls les fichiers qui correspondent au modèle spécifié sont copiés. Vous n'avez pas besoin de spécifier <code>--groupBy</code> et <code>--srcPattern</code> en même temps.</p> <p>Exemple : <code>--groupBy=.*subnetid.*([0-9]+-[0-9]+-[0-9]+-[0-9]+).</code>*</p>	Non

Option	Description	Obligatoire
<code>--targetSize=SIZE</code>	<p>Taille, en mébioctets (Mio), des fichiers à créer en fonction de l'option <code>--groupBy</code> . Cette valeur doit être un nombre entier. Lorsque cette option <code>--targetSize</code> est définie, S3 DistCp essaie de correspondre à cette taille ; la taille réelle des fichiers copiés peut être supérieure ou inférieure à cette valeur. Les travaux sont regroupés en fonction de la taille du fichier de données. Par conséquent, il est possible que la taille du fichier cible corresponde à la taille du fichier de données source.</p> <p>Si les fichiers concaténés par <code>--groupBy</code> sont plus grands que la valeur de <code>--targetSize</code> , ils sont divisés en fichiers plus petits, nommés de manière séquentielle avec une valeur numérique ajoutée à la fin de leur nom. Par exemple, un fichier concaténé <code>myfile.gz</code> peut être divisé en parties telles que <code>myfile0.gz</code> , <code>myfile1.gz</code> , etc.</p> <p>Exemple : <code>--targetSize=2</code></p>	Non
<code>--appendToLastFile</code>	<p>Spécifie le comportement de S3 DistCp lors de la copie vers des fichiers depuis Amazon S3 vers HDFS lesquels ils sont déjà présents. Il ajoute les données des nouveaux fichiers dans les fichiers existants. Si vous utilisez <code>--appendToLastFile</code> avec <code>--groupBy</code> , les nouvelles données sont ajoutées aux fichiers qui correspondent aux mêmes groupes. Cette option respecte également le comportement <code>--targetSize</code> lorsqu'il est utilisé avec <code>--groupBy</code> .</p>	Non

Option	Description	Obligatoire
<code>--outputCodec=CODEC</code>	<p>Spécifie le codec de compression à utiliser pour les fichiers copiés. Cela peut prendre les valeurs : <code>gzip</code>, <code>gz</code>, <code>lzo</code>, <code>snappy</code> ou <code>none</code>. Vous pouvez utiliser cette option, par exemple, pour convertir des fichiers d'entrée compressés avec Gzip en fichiers de sortie LZO compressés, ou pour décompresser les fichiers dans le cadre de l'opération de copie. Si vous choisissez un codec de sortie, le nom de fichier sera ajouté avec l'extension appropriée (par exemple, pour <code>gz</code> et <code>gzip</code>, l'extension est <code>.gz</code>). Si vous ne spécifiez pas de valeur pour <code>--outputCodec</code>, les fichiers sont copiés sans modification dans leur compression.</p> <p>Exemple : <code>--outputCodec=lzo</code></p>	Non
<code>--s3ServerSideEncryption</code>	<p>Garantit que les données cibles sont transférées SSL et chiffrées automatiquement dans Amazon S3 à l'aide d'une AWS clé côté service. Lorsque vous récupérez des données à l'aide de <code>S3DistCp</code>, les objets sont automatiquement déchiffrés. Si vous tentez de copier un objet non chiffré vers un compartiment Amazon S3 requérant le chiffrement, l'opération échoue. Pour plus d'informations, consultez Utilisation du chiffrement des données.</p> <p>Exemple : <code>--s3ServerSideEncryption</code></p>	Non

Option	Description	Obligatoire
<code>--deleteOnSuccess</code>	<p>Si l'opération de copie est réussie, cette option oblige S3 DistCp à supprimer les fichiers copiés de l'emplacement source. Cela est utile si vous copiez des fichiers de sortie, tels que les fichiers journaux, d'un emplacement à un autre comme une tâche planifiée, et que vous ne voulez pas copier deux fois les mêmes fichiers.</p> <p>Exemple : <code>--deleteOnSuccess</code></p>	Non
<code>--disableMultipartUpload</code>	<p>Désactive l'utilisation du téléchargement partitionné.</p> <p>Exemple : <code>--disableMultipartUpload</code></p>	Non
<code>--multipartUploadChunkSize=SIZE</code>	<p>Taille, en Mio, de chaque partie d'un chargement partitionné sur Amazon S3. S3 DistCp utilise le téléchargement partitionné lorsqu'il copie des données dont la <code>multipartUploadChunkSize</code> taille est supérieure à. Pour améliorer la performance au travail, vous pouvez augmenter la taille de chaque pièce. La taille par défaut est de 128 Mio.</p> <p>Exemple : <code>--multipartUploadChunkSize=1000</code></p>	Non
<code>--numberOfFiles</code>	<p>Ajoute aux fichiers de sortie des numéros séquentiels. Le compteur démarre à 0, à moins qu'une autre valeur soit spécifiée par <code>--startingIndex</code>.</p> <p>Exemple : <code>--numberOfFiles</code></p>	Non
<code>--startingIndex=INDEX</code>	<p>Utilisé avec <code>--numberOfFiles</code> pour spécifier le premier numéro de la séquence.</p> <p>Exemple : <code>--startingIndex=1</code></p>	Non

Option	Description	Obligatoire
<code>--outputManifest=FILENAME</code>	<p>Crée un fichier texte, compressé avec Gzip, qui contient une liste de tous les fichiers copiés par S3DistCp.</p> <p>Exemple : <code>--outputManifest=manifest-1.gz</code></p>	Non
<code>--previousManifest=PATH</code>	<p>Lit un fichier manifeste créé lors d'un appel précédent à S3 à DistCp l'aide de l'<code>--outputManifest</code> indicateur. Lorsque l'<code>--previousManifest</code> indicateur est défini, S3 DistCp exclut les fichiers répertoriés dans le manifeste de l'opération de copie. Si <code>--outputManifest</code> est spécifié avec <code>--previousManifest</code>, les fichiers répertoriés dans le manifeste précédent apparaissent également dans le nouveau fichier manifeste, même si les fichiers ne sont pas copiés.</p> <p>Exemple : <code>--previousManifest=/usr/bin/manifest-1.gz</code></p>	Non
<code>--requirePreviousManifest</code>	<p>Nécessite un manifeste créé lors d'un appel précédent à S3DistCp. Si la valeur spécifiée est false, aucune erreur n'est générée lorsqu'un manifeste précédent n'est pas spécifié. Par défaut, la valeur est true.</p>	Non
<code>--copyFromManifest</code>	<p>Inverse le comportement de S3 <code>--previousManifest</code> DistCp pour qu'il utilise le fichier manifeste spécifié comme liste de fichiers à copier, au lieu d'une liste de fichiers à exclure de la copie.</p> <p>Exemple : <code>--copyFromManifest --previousManifest=/usr/bin/manifest-1.gz</code></p>	Non

Option	Description	Obligatoire
<code>--s3Endpoint=ENDPOINT</code>	<p>Spécifie le point de terminaison Amazon S3 à utiliser lors du chargement d'un fichier. Cette option définit le point de terminaison pour la source et la destination. S'il n'est pas défini, le point de terminaison par défaut est <code>s3.amazonaws.com</code>.</p> <p>Pour obtenir la liste des points de terminaison Amazon S3, consultez Régions et points de terminaison.</p> <p>Exemple : <code>--s3Endpoint=s3.eu-west-1.amazonaws.com</code></p>	Non
<code>--storageClass=CLASS</code>	<p>La classe de stockage à utiliser lorsque la destination est Amazon S3. Les valeurs valides sont <code>STANDARD</code> et <code>REDUCED_REDUNDANCY</code>. Si cette option n'est pas spécifiée, S3 DistCp essaie de préserver la classe de stockage.</p> <p>Exemple : <code>--storageClass=STANDARD</code></p>	Non

Option	Description	Obligatoire
<code>--srcPrefixesFile=PATH</code>	<p>un fichier texte dans Amazon S3 (s3 ://), HDFS (hdfs ://) ou dans un système de fichiers local (file :/) qui contient une liste de <code>src</code> préfixes, un préfixe par ligne.</p> <p>S'il <code>srcPrefixesFile</code> est fourni, S3 ne DistCp listera pas le chemin <code>src</code>. Au lieu de cela, il génère une liste source en tant que résultat combiné de la liste de tous les préfixes spécifiés dans ce fichier. Le chemin d'accès relatif par rapport au chemin d'accès <code>src</code>, à la place de ces préfixes, sera utilisé pour générer les chemins de destination. Si <code>srcPattern</code> est également spécifié, il est appliqué aux résultats combinés de liste des préfixes sources pour filtrer encore plus les données d'entrée. Si <code>copyFromManifest</code> est utilisé, les objets dans le manifeste seront copiés et <code>srcPrefixesFile</code> sera ignoré.</p> <p>Exemple : <code>--srcPrefixesFile=PATH</code></p>	Non

Outre les options ci-dessus, S3 DistCp implémente l'[interface Tool](#), ce qui signifie qu'elle prend en charge les options génériques.

Ajouter S3 DistCp en tant qu'étape dans un cluster

Vous pouvez appeler S3 DistCp en l'ajoutant en tant qu'étape dans votre cluster. Les étapes peuvent être ajoutées à un cluster au lancement ou à un cluster en cours d'exécution à l'aide de la console CLI, ou API. Les exemples suivants illustrent l'ajout d'une DistCp étape S3 à un cluster en cours d'exécution. Pour plus d'informations sur l'ajout d'étapes à un cluster, consultez la section [Soumettre un travail à un cluster](#) dans le Amazon EMR Management Guide.

Pour ajouter une DistCp étape S3 à un cluster en cours d'exécution à l'aide du AWS CLI

Pour plus d'informations sur l'utilisation EMR des commandes Amazon dans le AWS CLI, consultez la [référence des AWS CLI commandes](#).

- Pour ajouter une étape à un cluster qui appelle S3DistCp, transmettez les paramètres qui spécifient comment S3 DistCp doit effectuer l'opération de copie en tant qu'arguments.

L'exemple suivant copie les journaux des démons de Amazon S3 dans `hdfs:///output`. Dans la commande suivante :

- `--cluster-id` spécifie le cluster
- `Jar` est l'emplacement du DistCp JAR fichier S3. Pour un exemple d'exécution d'une commande sur un cluster à l'aide de `command-runner.jar`, voir [Soumettre une JAR étape personnalisée pour exécuter un script ou une commande](#).
- `Args` est une liste séparée par des virgules des paires nom-valeur des options à transmettre à S3. DistCp Pour obtenir la liste complète des options disponibles, consultez [DistCp Options S3](#).

Pour ajouter une étape de DistCp copie S3 à un cluster en cours d'exécution, placez ce qui suit dans un JSON fichier enregistré dans Amazon S3 ou dans votre système de fichiers local, comme dans cet *myStep.json* exemple. Remplacez *j-3GYXXXXXX9I0K* avec votre identifiant de cluster et remplacez *mybucket* avec le nom de votre compartiment Amazon S3.

```
[
  {
    "Name": "S3DistCp step",
    "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXX9I0J/node/", "--dest=hdfs:///output", "--srcPattern=.*[a-zA-Z,]+"],
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR",
    "Jar": "command-runner.jar"
  }
]
```

```
aws emr add-steps --cluster-id j-3GYXXXXXX9I0K --steps file:///./myStep.json
```

Exemple Copiez les fichiers journaux depuis Amazon S3 vers HDFS

Cet exemple montre également comment copier des fichiers journaux stockés dans un compartiment Amazon S3 en HDFS ajoutant une étape à un cluster en cours d'exécution. Dans cet exemple, l'option `--srcPattern` est utilisée pour limiter les données copiées dans les journaux des démons.

Pour copier des fichiers journaux depuis Amazon S3 vers HDFS l'`--srcPattern` option, placez ce qui suit dans un JSON fichier enregistré dans Amazon S3 ou dans votre système de fichiers local, comme dans cet *myStep.json* exemple. Remplacez *j-3GYXXXXXX9I0K* avec votre identifiant de cluster et remplacez *mybucket* avec le nom de votre compartiment Amazon S3.

```
[
  {
    "Name": "S3DistCp step",
    "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXX9I0J/node/", "--dest=hdfs:///output", "--srcPattern=.*daemons.*-hadoop-.*"],
    "ActionOnFailure": "CONTINUE",
    "Type": "CUSTOM_JAR",
    "Jar": "command-runner.jar"
  }
]
```

Nettoyage après l'échec des DistCp tâches S3

Si S3 DistCp ne parvient pas à copier certains ou tous les fichiers spécifiés, l'étape de commande ou de cluster échoue et renvoie un code d'erreur différent de zéro. Dans ce cas, S3 DistCp ne nettoie pas les fichiers partiellement copiés. Vous devez les supprimer manuellement.

Les fichiers partiellement copiés sont enregistrés dans le HDFS tmp répertoire dans les sous-répertoires avec l'identifiant unique de la DistCp tâche S3. Cet identifiant figure dans la sortie standard de la tâche.

Par exemple, pour une DistCp tâche S3 avec l'ID4b1c37bb-91af-4391-aaf8-46a6067085a6, vous pouvez vous connecter au nœud principal du cluster et exécuter la commande suivante pour afficher les fichiers de sortie associés à la tâche.

```
hdfs dfs -ls /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6/output
```

La commande renvoie une liste de fichiers similaires à la liste suivante :

Found 8 items

```
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/_SUCCESS  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00000  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00001  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:02 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00002  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00003  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00004  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00005  
-rw-r--r-- 1 hadoop hadoop 0 2018-12-10 06:03 /tmp/4b1c37bb-91af-4391-  
aaf8-46a6067085a6/output/part-r-00006
```

Vous pouvez ensuite exécuter la commande suivante pour supprimer le répertoire et l'ensemble du contenu.

```
hdfs dfs rm -rf /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6
```

Exécuter des commandes et des scripts sur un cluster Amazon EMR

Cette rubrique explique comment exécuter une commande ou un script en tant qu'étape sur votre cluster. L'exécution d'une commande ou d'un script sous forme d'étape est l'une des nombreuses méthodes permettant de [soumettre du travail à un cluster](#). Elle est utile dans les situations suivantes :

- Lorsque vous n'avez pas accès SSH à votre cluster Amazon EMR
- Lorsque vous souhaitez exécuter une commande bash ou shell pour dépanner votre cluster

Vous pouvez exécuter un script soit lorsque vous créez un cluster, soit lorsque votre cluster est dans l'état WAITING. Pour exécuter un script avant le début du traitement de l'étape, utilisez plutôt une action d'amorçage. Pour plus d'informations sur les actions de démarrage, consultez la section [Création d'actions d'amorçage pour l'installation de logiciels supplémentaires](#) dans le Guide de gestion Amazon EMR.

Amazon EMR fournit les outils suivants pour vous aider à exécuter des scripts, des commandes et d'autres programmes intégrés au cluster. Vous pouvez appeler les deux outils à l'aide de la console de gestion Amazon EMR ou de l'AWS CLI.

`command-runner.jar`

Situé sur l'AMI Amazon EMR de votre cluster. Vous pouvez utiliser `command-runner.jar` pour exécuter des commandes sur votre cluster. Vous spécifiez `command-runner.jar` sans utiliser son chemin complet.

`script-runner.jar`

Hébergé sur Amazon S3 à `s3://<region>.elasticmapreduce/libs/script-runner/script-runner.jar` où `<region>` est la région dans laquelle réside votre cluster Amazon EMR. Vous pouvez utiliser `script-runner.jar` pour exécuter des scripts enregistrés localement ou sur Amazon S3 sur votre cluster. Vous devez spécifier l'URI complet de `script-runner.jar` lorsque vous soumettez une étape.

Soumettre une étape JAR personnalisée pour exécuter un script ou une commande

Les exemples AWS CLI suivants illustrent certains cas d'utilisation courants de `command-runner.jar` et `script-runner.jar` sur Amazon EMR.

Exemple : Exécution d'une commande sur un cluster à l'aide de **command-runner.jar**

Lorsque vous utilisez `command-runner.jar`, vous spécifiez des commandes, des options et des valeurs dans la liste d'arguments de votre étape.

L'exemple AWS CLI suivant soumet une étape à un cluster en cours d'exécution qui invoque `command-runner.jar`. La commande spécifiée dans la liste `Args` télécharge un script appelé *my-script.sh* depuis Amazon S3 dans le répertoire d'accueil de l'utilisateur `hadoop`. La commande modifie ensuite les autorisations du script et exécute *my-script.sh*.

Lorsque vous utilisez l'AWS CLI, les éléments de votre liste `Args` doivent être séparés par des virgules, sans espace entre les éléments de la liste. Par exemple, `Args=[example-command,example-option,"example option value"]` plutôt que `Args=[example-command, example-option, "example option value"]`.

```
aws emr add-steps \  
--cluster-id j-2AXXXXXXGAPLF \  
--steps Type=CUSTOM_JAR,Name="Download a script from S3, change its permissions, and  
run it",ActionOnFailure=CONTINUE,Jar=command-runner.jar,Args=[bash,-c,"aws s3 cp s3://  
EXAMPLE-DOC-BUCKET/my-script.sh /home/hadoop; chmod u+x /home/hadoop/my-script.sh; cd /  
home/hadoop; ./my-script.sh"]
```

Exemple : Exécution d'un script sur un cluster à l'aide de **script-runner.jar**

Lorsque vous utilisez `script-runner.jar`, vous spécifiez le script que vous souhaitez exécuter dans la liste des arguments de votre étape.

L'exemple AWS CLI suivant soumet une étape à un cluster en cours d'exécution qui invoque `script-runner.jar`. Dans ce cas, le script appelé *my-script.sh* est stocké sur Amazon S3. Vous pouvez également spécifier des scripts locaux qui sont stockés sur le nœud principal de votre cluster.

```
aws emr add-steps \  
--cluster-id j-2AXXXXXXGAPLF \  

```

```
--steps Type=CUSTOM_JAR,Name="Run a script from S3 with script-
runner.jar",ActionOnFailure=CONTINUE,Jar=s3://us-west-2.elasticmapreduce/libs/script-
runner/script-runner.jar,Args=[s3://EXAMPLE-DOC-BUCKET/my-script.sh]
```

Autres moyens d'utiliser `command-runner.jar`

Vous pouvez également utiliser `command-runner.jar` pour soumettre du travail à un cluster à l'aide d'outils tels que `spark-submit` ou `hadoop-streaming`. Lorsque vous lancez une application à l'aide de `command-runner.jar`, vous spécifiez `CUSTOM_JAR` comme type d'étape au lieu d'utiliser une valeur telle que `SPARK`, `STREAMING` ou `PIG`. La disponibilité des outils varie en fonction des applications que vous avez installées sur le cluster.

L'exemple de commande suivant utilise `command-runner.jar` pour soumettre une étape à l'aide de `spark-submit`. La liste `Args` indique `spark-submit` comme commande, suivie de l'URI Amazon S3 de l'application Spark `my-app.py` avec des arguments et des valeurs.

```
aws emr add-steps \
--cluster-id j-2AXXXXXXGAPLF \
--steps Type=CUSTOM_JAR,Name="Run spark-submit using command-
runner.jar",ActionOnFailure=CONTINUE,Jar=command-runner.jar,Args=[spark-submit,S3://
DOC-EXAMPLE-BUCKET/my-app.py,ArgName1,ArgValue1,ArgName2,ArgValue2]
```

Le tableau suivant identifie les outils supplémentaires que vous pouvez exécuter à l'aide de `command-runner.jar`.

Nom de l'outil	Description
<code>hadoop-streaming</code>	Soumet un programme de streaming Hadoop. Dans la console et certains kits SDK, il s'agit d'une étape de streaming.
<code>hive-script</code>	Exécute un script Hive. Dans la console et les kits SDK, il s'agit d'une étape Hive.
<code>pig-script</code>	Exécute un script Pig. Dans la console et les kits SDK, il s'agit d'une étape Pig.
<code>spark-submit</code>	Exécute une application Spark. Dans la console, il s'agit d'une étape Spark.

Nom de l'outil	Description
hadoop-lzo	Exécute l' indexeur Hadoop LZO sur un répertoire.
s3-dist-cp	Effectuez la copie distribuée de grandes quantités de données d'Amazon S3 dans HDFS. Pour de plus amples informations, veuillez consulter S3 DistCp (s3-dist-cp) .

Glossaire AWS

Pour connaître la terminologie la plus récente d'AWS, consultez le [Glossaire AWS](#) dans la Référence Glossaire AWS.

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.